# Gretl Command Reference

Gnu Regression, Econometrics and Time-series Library

Allin Cottrell
Department of Economics
Wake Forest University

Riccardo "Jack" Lucchetti
Dipartimento di Economia
Università Politecnica delle Marche

April, 2012

# Contents

Contents                                                                                    iii

Contents

Contents

Contents

# Gretl commands

## 1.1 Introduction

The commands defined below may be executed interactively in the command-line client program or in the console window of the GUI program. They may also be placed in a "script" or batch file for non-interactive execution.

The following notational conventions are used below:

- A `typewriter font` is used for material that you would type directly, and also for internal names of variables.

- Terms in a *slanted font* are place-holders: you should substitute some specific replacement. For example, you might type `income` in place of the generic *xvar*.

- The construction [ *arg* ] means that the argument *arg* is optional: you may supply it or not (but in any case don't type the brackets).

- The phrase "estimation command" means a command that generates estimates for a given model, for example `ols`, `ar` or `wls`.

In general, each line of a command script should contain one and only one complete gretl command. There are, however, two means of continuing a long command from one line of input to another. First, if the last non-space character on a line is a backslash, this is taken as an indication that the command is continued on the following line. In addition, if the comma is a valid character in a given command (for instance, as a separator between function arguments, or as punctuation in the command `printf`) then a trailing comma also indicates continuation. To emphasize the point: a backslash may be inserted "arbitrarily" to indicate continuation, but a comma works in this capacity only if it is syntactically valid as part of the command.

## 1.2 Commands

**add**

| | |
|---|---|
| Argument: | *varlist* |
| Options: | `--lm` (do an LM test, OLS only) |
| | `--quiet` (print only the basic test result) |
| | `--silent` (don't print anything) |
| | `--vcv` (print covariance matrix for augmented model) |
| | `--both` (IV estimation only, see below) |
| Examples: | `add 5 7 9` |
| | `add xx yy zz --quiet` |

Must be invoked after an estimation command. Performs a joint test for the addition of the specified variables to the last model, the results of which may be retrieved using the accessors `$test` and `$pvalue`.

By default an augmented version of the original model is estimated, including the variables in *varlist*. The test is a Wald test on the augmented model, which replaces the original as the "current model" for the purposes of, for example, retrieving the residuals as `$uhat` or doing further tests.

Alternatively, given the `--lm` option (available only for the models estimated via OLS), an LM test is performed. An auxiliary regression is run in which the dependent variable is the residual from the last model and the independent variables are those from the last model plus *varlist*. Under the null hypothesis that the added variables have no additional explanatory power, the sample size times the unadjusted R-squared from this regression is distributed as chi-square with degrees of freedom equal to the number of added regressors. In this case the original model is not replaced.

The `--both` option is specific to two-stage least squares: it specifies that the new variables should be added both to the list of regressors and the list of instruments, the default in this case being to add to the regressors only.

Menu path: Model window, /Tests/Add variables

**adf**

| | |
|---|---|
| Arguments: | *order varlist* |
| Options: | `--nc` (test without a constant) |
| | `--c` (with constant only) |
| | `--ct` (with constant and trend) |
| | `--ctt` (with constant, trend and trend squared) |
| | `--seasonals` (include seasonal dummy variables) |
| | `--gls` (de-mean or de-trend using GLS) |
| | `--verbose` (print regression results) |
| | `--quiet` (suppress printing of results) |
| | `--difference` (use first difference of variable) |
| | `--test-down` (automatic lag order) |
| Examples: | `adf 0 y` |
| | `adf 2 y --nc --c --ct` |
| | `adf 12 y --c --test-down` |
| | See also `jgm-1996.inp` |

The options shown above and the discussion which follows pertain to the use of the `adf` command with regular time series data. For use of this command with panel data please see below.

Computes a set of Dickey–Fuller tests on each of the the listed variables, the null hypothesis being that the variable in question has a unit root. (But if the `--difference` flag is given, the first difference of the variable is taken prior to testing, and the discussion below must be taken as referring to the transformed variable.)

By default, two variants of the test are shown: one based on a regression containing a constant and one using a constant and linear trend. You can control the variants that are presented by specifying one or more of the option flags.

In all cases the dependent variable is the first difference of the specified variable, $y$, and the key independent variable is the first lag of $y$. The model is constructed so that the coefficient on lagged $y$ equals the root in question minus 1. For example, the model with a constant may be written as

$$(1 - L)y_t = \beta_0 + (\alpha - 1)y_{t-1} + \epsilon_t$$

Under the null hypothesis of a unit root the coefficient on lagged $y$ equals zero; under the alternative that $y$ is stationary this coefficient is negative.

If the lag order, $k$, is greater than 0, then $k$ lags of the dependent variable are included on the right-hand side of the test regressions. If the order is given as $-1$, $k$ is set following the recommendation of Schwert (1989), namely $12(T/100)^{0.25}$, where $T$ is the sample size. In either case, however, if the `--test-down` option is given then $k$ is taken as the *maximum* lag and the actual lag order used is obtained by testing down, using this algorithm:

1. Estimate the Dickey–Fuller regression with $k$ lags of the dependent variable.

2. Is the last lag significant? If so, execute the test with lag order $k$. Otherwise, let $k = k - 1$;
   if $k$ equals 0, execute the test with lag order 0, else go to step 1.

In the context of step 2 above, "significant" means that the $t$-statistic for the last lag has an asymptotic two-sided $p$-value, against the normal distribution, of 0.10 or less.

The `--gls` option can be used in conjunction with one or other of the flags `--c` and `--ct` (the model with constant, or model with constant and trend). The effect of this option is that the de-meaning or de-trending of the variable to be tested is done using the GLS procedure suggested by Elliott *et al.* (1996), which gives a test of greater power than the standard Dickey–Fuller approach. This option is not compatible with `--nc`, `--ctt` or `--seasonals`.

*P*-values for the Dickey–Fuller tests are based on MacKinnon (1996). The relevant code is included by kind permission of the author. In the case of the test with linear trend using GLS these *P*-values are not applicable; critical values from Table 1 in Elliott *et al.* (1996) are shown instead.

*Panel data*

When the `adf` command is used with panel data, to produce a panel unit root test, the applicable options and the results shown are somewhat different.

First, while you may give a list of variables for testing in the regular time-series case, with panel data only one variable may be tested per command. Second, the options governing the inclusion of deterministic terms become mutually exclusive: you must choose between no-constant, constant only, and constant plus trend; the default is constant only. In addition, the `--seasonals` option is not available. Third, the `--verbose` option has a different meaning: it produces a brief account of the test for each individual time series (the default being to show only the overall result).

The overall test (null hypothesis: the series in question has a unit root for all the panel units) is calculated in one or both of two ways: using the method of Im *et al.* (2003) or that of Choi (2001).

Menu path: /Variable/Unit root tests/Augmented Dickey-Fuller test

**anova**

  Arguments:    *response treatment* [ *block* ]
  Option:       `--quiet` (don't print results)

Analysis of Variance: *response* is a series measuring some effect of interest and *treatment* must be a discrete variable that codes for two or more types of treatment (or non-treatment). For two-way ANOVA, the *block* variable (which should also be discrete) codes for the values of some control variable.

Unless the `--quiet` option is given, this command prints a table showing the sums of squares and mean squares along with an $F$-test. The $F$-test and its P-value can be retrieved using the accessors `$test` and `$pvalue` respectively.

The null hypothesis for the $F$-test is that the mean response is invariant with respect to the treatment type, or in words that the treatment has no effect. Strictly speaking, the test is valid only if the variance of the response is the same for all treatment types.

Note that the results shown by this command are in fact a subset of the information given by the following procedure, which is easily implemented in gretl. Create a set of dummy variables coding for all but one of the treatment types. For two-way ANOVA, in addition create a set of dummies coding for all but one of the "blocks". Then regress *response* on a constant and the dummies using ols. For a one-way design the ANOVA table is printed via the `--anova` option to ols. In the two-way case the relevant $F$-test is found by using the omit command. For example (assuming y is the response, `xt` codes for the treatment, and `xb` codes for blocks):

```
# one-way
list dxt = dummify(xt)
```

```
ols y 0 dxt --anova
# two-way
list dxb = dummify(xb)
ols y 0 dxt dxb
# test joint significance of dxt
omit dxt --quiet
```

Menu path: /Model/Other linear models/ANOVA

## append

  Argument:    *filename*
  Option:      `--time-series` (see below)

Opens a data file and appends the content to the current dataset, if the new data are compatible. The program will try to detect the format of the data file (native, plain text, CSV, Gnumeric, Excel, etc.).

The appended data may take the form of either additional observations on variables already present in the dataset, or new variables. in the case of adding variables, compatibility requires either (a) that the number of observations for the new data equals that for the current data, or (b) that the new data carries clear observation information so that gretl can work out how to place the values.

A special feature is supported for appending to a panel dataset. Let $n$ denote the number of cross-sectional units in the panel, $T$ denote the number of time periods, and $m$ denote the number of observations for the new data. If $m = n$ the new data are taken to be time-invariant, and are copied into place for each time period. On the other hand, if $m = T$ the data are treated as non-varying across the panel units, and are copied into place for each unit. If the panel is "square", and $m$ equals both $n$ and $T$, an ambiguity arises. The default in this case is to treat the new data as time-invariant, but you can force gretl to treat the new data as time series via the `--time-series` option. (This option is ignored in all other cases.)

Menu path: /File/Append data

## ar

  Arguments:   *lags* ; *depvar indepvars*
  Option:      `--vcv` (print covariance matrix)
  Example:     `ar 1 3 4 ; y 0 x1 x2 x3`

Computes parameter estimates using the generalized Cochrane–Orcutt iterative procedure; see Section 9.5 of Ramanathan (2002). Iteration is terminated when successive error sums of squares do not differ by more than 0.005 percent or after 20 iterations.

*lags* is a list of lags in the residuals, terminated by a semicolon. In the above example, the error term is specified as

$$u_t = \rho_1 u_{t-1} + \rho_3 u_{t-3} + \rho_4 u_{t-4} + e_t$$

Menu path: /Model/Time series/Autoregressive estimation

## ar1

  Arguments:   *depvar indepvars*
  Options:     `--hilu` (use Hildreth–Lu procedure)
               `--pwe` (use Prais–Winsten estimator)
               `--vcv` (print covariance matrix)
               `--no-corc` (do not fine-tune results with Cochrane-Orcutt)
  Examples:    `ar1 1 0 2 4 6 7`
               `ar1 y 0 xlist --hilu --no-corc`
               `ar1 y 0 xlist --pwe`

Computes feasible GLS estimates for a model in which the error term is assumed to follow a first-order autoregressive process.

The default method is the Cochrane–Orcutt iterative procedure; see for example section 9.4 of Ramanathan (2002). Iteration is terminated when successive estimates of the autocorrelation coefficient do not differ by more than 0.001 or after 20 iterations.

If the `--hilu` option is given, the Hildreth–Lu search procedure is used. The results are then fine-tuned using the Cochrane–Orcutt method, unless the `--no-corc` flag is specified. (The latter option is ignored if `--hilu` is not specified.)

If the `--pwe` option is given, the Prais–Winsten estimator is used. This involves an an iteration similar to Cochrane–Orcutt; the difference is that while Cochrane–Orcutt discards the first observation, Prais–Winsten makes use of it. See, for example, Chapter 13 of Greene (2000) for details.

Menu path: /Model/Time series/Cochrane-Orcutt

Menu path: /Model/Time series/Hildreth-Lu

Menu path: /Model/Time series/Prais-Winsten

### arbond

| | |
|---|---|
| Argument: | *p* [ *q* ] ; *depvar indepvars* [ ; *instruments* ] |
| Options: | `--quiet` (don't show estimated model) |
| | `--vcv` (print covariance matrix) |
| | `--two-step` (perform 2-step GMM estimation) |
| | `--time-dummies` (add time dummy variables) |
| | `--asymptotic` (uncorrected asymptotic standard errors) |
| Examples: | `arbond 2 ; y Dx1 Dx2` |
| | `arbond 2 5 ; y Dx1 Dx2 ; Dx1` |
| | `arbond 1 ; y Dx1 Dx2 ; Dx1 GMM(x2,2,3)` |
| | See also `arbond91.inp` |

Carries out estimation of dynamic panel data models (that is, panel models including one or more lags of the dependent variable) using the GMM-DIF method set out by Arellano and Bond (1991). Please see dpanel for an updated and more flexible version of this command which handles GMM-SYS as well as GMM-DIF.

The parameter $p$ represents the order of the autoregression for the dependent variable. The optional parameter $q$ indicates the maximum lag of the level of the dependent variable to be used as an instrument. If this argument is omitted, or given as 0, all available lags are used.

The dependent variable should be given in levels form; it will be automatically differenced (since this estimator uses differencing to cancel out the individual effects). The independent variables are not automatically differenced; if you want to use differences (which will generally be the case for ordinary quantitative variables, though perhaps not for, say, time dummy variables) you should create the differences first then specify these as the regressors.

The last (optional) field in the command is for specifying instruments. If no instruments are given, it is assumed that all the independent variables are strictly exogenous. If you specify any instruments, you should include in the list any strictly exogenous independent variables. For predetermined regressors, you can use the GMM function to include a specified range of lags in block-diagonal fashion. This is illustrated in the third example above. The first argument to GMM is the name of the variable in question, the second is the minimum lag to be used as an instrument, and the third is the maximum lag. If the third argument is given as 0, all available lags are used.

By default the results of 1-step estimation are reported (with robust standard errors). You may select 2-step estimation as an option. In both cases tests for autocorrelation of orders 1 and 2 are provided, as well as the Sargan overidentification test and a Wald test for the joint significance of the regressors. Note that in this differenced model first-order autocorrelation is not

a threat to the validity of the model, but second-order autocorrelation violates the maintained statistical assumptions.

In the case of 2-step estimation, standard errors are by default computed using the finite-sample correction suggested by Windmeijer (2005). The standard asymptotic standard errors associated with the 2-step estimator are generally reckoned to be an unreliable guide to inference, but if for some reason you want to see them you can use the `--asymptotic` option to turn off the Windmeijer correction.

If the `--time-dummies` option is given, a set of time dummy variables is added to the specified regressors. The number of dummies is one less than the maximum number of periods used in estimation, to avoid perfect collinearity with the constant. The dummies are entered in levels; if you wish to use time dummies in first-differenced form, you will have to define and add these variables manually.

### arch

|  |  |
|---|---|
| Arguments: | *order depvar indepvars* |
| Example: | `arch 4 y 0 x1 x2 x3` |

This command is retained at present for backward compatibility, but you are better off using the maximum likelihood estimator offered by the garch command; for a plain ARCH model, set the first GARCH parameter to 0.

Estimates the given model specification allowing for ARCH (Autoregressive Conditional Heteroskedasticity). The model is first estimated via OLS, then an auxiliary regression is run, in which the squared residual from the first stage is regressed on its own lagged values. The final step is weighted least squares estimation, using as weights the reciprocals of the fitted error variances from the auxiliary regression. (If the predicted variance of any observation in the auxiliary regression is not positive, then the corresponding squared residual is used instead).

The `alpha` values displayed below the coefficients are the estimated parameters of the ARCH process from the auxiliary regression.

See also garch and modtest (the `--arch` option).

Menu path: /Model/Time series/ARCH

### arima

|  |  |
|---|---|
| Arguments: | *p d q* [ ; *P D Q* ] ; *depvar* [ *indepvars* ] |
| Options: | `--verbose` (print details of iterations) |
|  | `--vcv` (print covariance matrix) |
|  | `--hessian` (see below) |
|  | `--opg` (see below) |
|  | `--nc` (do not include a constant) |
|  | `--conditional` (use conditional maximum likelihood) |
|  | `--x-12-arima` (use X-12-ARIMA for estimation) |
|  | `--lbfgs` (use L-BFGS-B maximizer) |
|  | `--y-diff-only` (ARIMAX special, see below) |
|  | `--save-ehat` (see below) |
| Examples: | `arima 1 0 2 ; y` |
|  | `arima 2 0 2 ; y 0 x1 x2 --verbose` |
|  | `arima 0 1 1 ; 0 1 1 ; y --nc` |

If no *indepvars* list is given, estimates a univariate ARIMA (Autoregressive, Integrated, Moving Average) model. The values *p*, *d* and *q* represent the autoregressive (AR) order, the differencing order, and the moving average (MA) order respectively. These values may be given in numerical form, or as the names of pre-existing scalar variables. A *d* value of 1, for instance, means that the first difference of the dependent variable should be taken before estimating the ARMA

parameters.

If you wish to include only specific AR or MA lags in the model (as opposed to all lags up to a given order) you can substitute for $p$ and/or $q$ either (a) the name of a pre-defined matrix containing a set of integer values or (b) an expression such as {1 4}; that is, a set of lags separated by spaces and enclosed in braces.

The optional integer values $P$, $D$ and $Q$ represent the seasonal AR, order for seasonal differencing and seasonal MA order respectively. These are applicable only if the data have a frequency greater than 1 (for example, quarterly or monthly data). These orders must be given in numerical form or as scalar variables.

In the univariate case the default is to include an intercept in the model but this can be suppressed with the --nc flag. If *indepvars* are added, the model becomes ARMAX; in this case the constant should be included explicitly if you want an intercept (as in the second example above).

An alternative form of syntax is available for this command: if you do not want to apply differencing (either seasonal or non-seasonal), you may omit the $d$ and $D$ fields altogether, rather than explicitly entering 0. In addition, arma is a synonym or alias for arima. Thus for example the following command is a valid way to specify an ARMA(2, 1) model:

        arma 2 1 ; y

The default is to use the "native" gretl ARMA functionality, with estimation by exact ML using the Kalman filter; estimation via conditional ML is available as an option. (If X-12-ARIMA is installed you have the option of using it instead of native code.) For details regarding these options, please see the *Gretl User's Guide.*

When the native exact ML code is used, estimated standard errors are by default based on a numerical approximation to the (negative inverse of) the Hessian, with a fallback to the outer product of the gradient (OPG) if calculation of the numerical Hessian should fail. Two (mutually exclusive) option flags can be used to force the issue: the --opg option forces use of the OPG method, with no attempt to compute the Hessian, while the --hessian flag disables the fallback to OPG. Note that failure of the numerical Hessian computation is generally an indicator of a misspecified model.

The option --lbfgs is specific to estimation using native ARMA code and exact ML: it calls for use of the "limited memory" L-BFGS-B algorithm in place of the regular BFGS maximizer. This may help in some instances where convergence is difficult to achieve.

The option --y-diff-only is specific to estimation of ARIMAX models (models with a non-zero order of integration and including exogenous regressors), and applies only when gretl's native exact ML is used. For such models the default behavior is to difference both the dependent variable and the regressors, but when this option is specified only the dependent variable is differenced, the regressors remaining in level form.

The option --save-ehat is applicable only when using native exact ML estimation. The effect is to make available a vector holding the optimal estimate as of period $t$ of the $t$-dated disturbance or innovation: this can be retrieved via the accessor $ehat. These values differ from the residual series ($uhat), which holds the one-step ahead forecast errors.

The AIC value given in connection with ARIMA models is calculated according to the definition used in X-12-ARIMA, namely

$$\text{AIC} = -2\ell + 2k$$

where $\ell$ is the log-likelihood and $k$ is the total number of parameters estimated. Note that X-12-ARIMA does not produce information criteria such as AIC when estimation is by conditional ML.

The AR and MA roots shown in connection with ARMA estimation are based on the following representation of an ARMA($p, q$) process:

$$(1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)Y = c + (1 + \theta_1 L + \theta_2 L^2 + \cdots + \theta_q L^q)\varepsilon_t$$

The AR roots are therefore the solutions to

$$1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p L^p = 0$$

and stability requires that these roots lie outside the unit circle.

The "frequency" figure printed in connection with the AR and MA roots is the $\lambda$ value that solves $z = re^{i2\pi\lambda}$, where $z$ is the root in question and $r$ is its modulus.

Menu path: /Model/Time series/ARIMA

Other access: Main window pop-up menu (single selection)

### biprobit

| | |
|---|---|
| Arguments: | *depvar1 depvar2 indepvars1* [ ; *indepvars2* ] |
| Options: | `--vcv` (print covariance matrix) |
| | `--robust` (robust standard errors) |
| | `--opg` (see below) |
| | `--save-xbeta` (see below) |
| | `--verbose` (print extra information) |
| Examples: | `biprobit y1 y2 0 x1 x2` |
| | `biprobit y1 y2 0 x11 x12 ; 0 x21 x22` |
| | See also `biprobit.inp` |

Estimates a bivariate probit model, using the Newton–Raphson method to maximize the likelihood.

The argument list starts with the two (binary) dependent variables, followed by a list of regressors. If a second list is given, separated by a semicolon, this is interpreted as a set of regressors specific to the second equation, with *indepvars1* being specific to the first equation; otherwise *indepvars1* is taken to represent a common set of regressors.

By default, standard errors are computed using a numerical approximation to the Hessian at convergence. But if the `--opg` option is given the covariance matrix is based on the Outer Product of the Gradient (OPG), or if the `--robust` option is given QML standard errors are calculated, using a "sandwich" of the inverse of the Hessian and the OPG.

After successful estimation, the accessor `$uhat` retrieves a matrix with two columns holding the generalized residuals for the two equations; that is, the expected values of the disturbances conditional on the observed outcomes and covariates. By default `$yhat` retrieves a matrix with four columns, holding the estimated probabilities of the four possible joint outcomes for ($y_1$, $y_2$), in the order (1,1), (1,0), (0,1), (0,0). Alternatively, if the option `--save-xbeta` is given, `$yhat` has two columns and holds the values of the index functions for the respective equations.

The output includes a likelihood ratio test of the null hypothesis that the disturbances in the two equations are uncorrelated.

### boxplot

| | |
|---|---|
| Argument: | *varlist* |
| Options: | `--notches` (show 90 percent interval for median) |
| | `--factorized` (see below) |
| | `--panel` (see below) |
| | `--output=`*filename* (send output to specified file) |

These plots display the distribution of a variable. The central box encloses the middle 50 percent of the data, i.e. it is bounded by the first and third quartiles. The "whiskers" extend to the minimum and maximum values. A line is drawn across the box at the median. A "+" sign is used to indicate the mean. If the option of showing a confidence interval for the median is selected, this is computed via the bootstrap method and shown in the form of dashed horizontal lines above and/or below the median.

The `--factorized` option allows you to examine the distribution of a chosen variable conditional on the value of some discrete factor.  For example, if a data set contains wages and a gender dummy variable you can select the wage variable as the target and gender as the factor, to see side-by-side boxplots of male and female wages, as in

```
boxplot wage gender --factorized
```

Note that in this case you must specify exactly two variables, with the factor given second.

If the current data set is a panel, and just one variable is specified, the `--panel` option produces a series of side-by-side boxplots, one for each panel "unit" or group.

In interactive mode the result is displayed immediately. In batch mode the default behavior is that a gnuplot command file is written in the user's working directory, with a name on the pattern `gpttmpN.plt`, starting with N = 01. The actual plots may be generated later using gnuplot (under MS Windows, wgnuplot). This behavior can be modified by use of the `--output=`*filename* option. For details, please see the gnuplot command.

Menu path: /View/Graph specified vars/Boxplots

### break

Break out of a loop. This command can be used only within a loop; it causes command execution to break out of the current (innermost) loop. See also loop.

### catch

  Syntax:   `catch` *command*

This is not a command in its own right but can be used as a prefix to most regular commands: the effect is to prevent termination of a script if an error occurs in executing the command. If an error does occur, this is registered in an internal error code which can be accessed as `$error` (a zero value indicates success).  The value of `$error` should always be checked immediately after using `catch`, and appropriate action taken if the command failed.

The `catch` keyword cannot be used before `if`, `elif` or `endif`.

### chow

| Variants: | `chow` *obs* |
| --- | --- |
| | `chow` *dummyvar* `--dummy` |
| Options: | `--dummy` (use a pre-existing dummy variable) |
| | `--quiet` (don't print estimates for augmented model) |
| Examples: | `chow 25` |
| | `chow 1988:1` |
| | `chow female --dummy` |

Must follow an OLS regression.  If an observation number or date is given, provides a test for the null hypothesis of no structural break at the given split point. The procedure is to create a dummy variable which equals 1 from the split point specified by *obs* to the end of the sample, 0 otherwise, and also interaction terms between this dummy and the original regressors. If a dummy variable is given, tests the null hypothesis of structural homogeneity with respect to that dummy. Again, interaction terms are added. In either case an augmented regression is run including the additional terms.

By default an $F$ statistic is calculated, taking the augmented regression as the unrestricted model and the original as the restricted. But if the original model used a robust estimator for the covariance matrix, the test statistic is a Wald chi-square value based on a robust estimator of the covariance matrix for the augmented regression.

Menu path: Model window, /Tests/Chow test

**clear**

Options:   --dataset (clear dataset only)

--other (clear everything but the dataset)

With no options, clears all saved objects, including the current dataset if any, out of memory. Note that opening a new dataset, or using the nulldata command to create an empty dataset, also has this effect, so use of clear is not usually necessary.

If the --dataset option is given, then only the dataset is cleared; other saved objects such as named matrices and scalars are preserved.

If the --other option is given, then the current dataset, if any, is preserved but all other saved objects are destroyed.

**coeffsum**

Argument:   *varlist*

Example:    coeffsum xt xt_1 xr_2

restrict.inp

Must follow a regression. Calculates the sum of the coefficients on the variables in *varlist*. Prints this sum along with its standard error and the p-value for the null hypothesis that the sum is zero.

Note the difference between this and omit, which tests the null hypothesis that the coefficients on a specified subset of independent variables are *all* equal to zero.

Menu path: Model window, /Tests/Sum of coefficients

**coint**

Arguments:   *order depvar indepvars*

Options:     --nc (do not include a constant)

--ct (include constant and trend)

--ctt (include constant and quadratic trend)

--skip-df (no DF tests on individual variables)

--test-down (automatic lag order)

--verbose (print extra details of regressions)

Examples:    coint 4 y x1 x2

coint 0 y x1 x2 --ct --skip-df

The Engle and Granger (1987) cointegration test. The default procedure is: (1) carry out Dickey–Fuller tests on the null hypothesis that each of the variables listed has a unit root; (2) estimate the cointegrating regression; and (3) run a DF test on the residuals from the cointegrating regression. If the --skip-df flag is given, step (1) is omitted.

If the specified lag order is positive all the Dickey–Fuller tests use that order, with this qualification: if the --test-down option is given, the given value is taken as the maximum and the actual lag order used in each case is obtained by testing down. See the adf command for details of this procedure.

By default, the cointegrating regression contains a constant. If you wish to suppress the constant, add the --nc flag. If you wish to augment the list of deterministic terms in the cointegrating regression with a linear or quadratic trend, add the --ct or --ctt flag. These option flags are mutually exclusive.

*P*-values for this test are based on MacKinnon (1996). The relevant code is included by kind permission of the author.

Menu path: /Model/Time series/Cointegration test/Engle-Granger

**coint2**

| | |
|---|---|
| Arguments: | *order ylist* [ ; *xlist* ] [ ; *rxlist* ] |
| Options: | `--nc` (no constant) |
| | `--rc` (restricted constant) |
| | `--uc` (unrestricted constant) |
| | `--crt` (constant and restricted trend) |
| | `--ct` (constant and unrestricted trend) |
| | `--seasonals` (include centered seasonal dummies) |
| | `--asy` (record asymptotic p-values) |
| | `--quiet` (print just the tests) |
| | `--silent` (don't print anything) |
| | `--verbose` (print details of auxiliary regressions) |
| Examples: | `coint2 2 y x` |
| | `coint2 4 y x1 x2 --verbose` |
| | `coint2 3 y x1 x2 --rc` |

Carries out the Johansen test for cointegration among the variables in *ylist* for the given lag order. For details of this test see the *Gretl User's Guide* or Hamilton (1994), Chapter 20. P-values are computed via Doornik's gamma approximation (Doornik, 1998). Two sets of p-values are shown for the trace test, straight asymptotic values and values adjusted for the sample size. By default the `$pvalue` accessor gets the adjusted variant, but the `--asy` flag may be used to record the asymptotic values instead.

The inclusion of deterministic terms in the model is controlled by the option flags. The default if no option is specified is to include an "unrestricted constant", which allows for the presence of a non-zero intercept in the cointegrating relations as well as a trend in the levels of the endogenous variables. In the literature stemming from the work of Johansen (see for example his 1995 book) this is often referred to as "case 3". The first four options given above, which are mutually exclusive, produce cases 1, 2, 4 and 5 respectively. The meaning of these cases and the criteria for selecting a case are explained in the *Gretl User's Guide*.

The optional lists *xlist* and *rxlist* allow you to control for specified exogenous variables: these enter the system either unrestrictedly (*xlist*) or restricted to the cointegration space (*rxlist*). These lists are separated from *ylist* and from each other by semicolons.

The `--seasonals` option, which may be combined with any of the other options, specifies the inclusion of a set of centered seasonal dummy variables. This option is available only for quarterly or monthly data.

The following table is offered as a guide to the interpretation of the results shown for the test, for the 3-variable case. $H_0$ denotes the null hypothesis, $H_1$ the alternative hypothesis, and $c$ the number of cointegrating relations.

| | Trace test | | $\lambda$-max test | |
|---|---|---|---|---|
| Rank | $H_0$ | $H_1$ | $H_0$ | $H_1$ |
| 0 | $c = 0$ | $c = 3$ | $c = 0$ | $c = 1$ |
| 1 | $c = 1$ | $c = 3$ | $c = 1$ | $c = 2$ |
| 2 | $c = 2$ | $c = 3$ | $c = 2$ | $c = 3$ |

See also the vecm command.

Menu path: /Model/Time series/Cointegration test/Johansen

**corr**

| Argument: | [ *varlist* ] |
|---|---|
| Options: | `--uniform` (ensure uniform sample) |
| | `--spearman` (Spearman's rho) |
| | `--kendall` (Kendall's tau) |
| | `--verbose` (print rankings) |
| Examples: | `corr y x1 x2 x3` |
| | `corr ylist --uniform` |
| | `corr x y --spearman` |

By default, prints the pairwise correlation coefficients (Pearson's product-moment correlation) for the variables in *varlist*, or for all variables in the data set if *varlist* is not given. The standard behavior is to use all available observations for computing each pairwise coefficient, but if the `--uniform` option is given the sample is limited (if necessary) so that the same set of observations is used for all the coefficients. This option has an effect only if there are differing numbers of missing values for the variables used.

The (mutually exclusive) options `--spearman` and `--kendall` produce, respectively, Spearman's rank correlation rho and Kendall's rank correlation tau in place of the default Pearson coefficient. When either of these options is given, *varlist* should contain just two variables.

When a rank correlation is computed, the `--verbose` option can be used to print the original and ranked data (otherwise this option is ignored).

Menu path: /View/Correlation matrix

Other access: Main window pop-up menu (multiple selection)

**corrgm**

| Arguments: | *series* [ *order* ] |
|---|---|
| Option: | `--quiet` (don't draw ACF graph) |
| Example: | `corrgm x 12` |

Prints the values of the autocorrelation function for *series*, which may be specified by name or number. The values are defined as $\hat{\rho}(u_t, u_{t-s})$ where $u_t$ is the $t$th observation of the variable $u$ and $s$ is the number of lags.

The partial autocorrelations (calculated using the Durbin–Levinson algorithm) are also shown: these are net of the effects of intervening lags. In addition the Ljung–Box $Q$ statistic is printed. This may be used to test the null hypothesis that the series is "white noise"; it is asymptotically distributed as chi-square with degrees of freedom equal to the number of lags used. Unless the `--quiet` option is given, a plot of the correlogram is printed.

If a *order* value is specified the length of the correlogram is limited to at most that number of lags, otherwise the length is determined automatically, as a function of the frequency of the data and the number of observations.

Menu path: /Variable/Correlogram

Other access: Main window pop-up menu (single selection)

**cusum**

| Options: | `--squares` (perform the CUSUMSQ test) |
|---|---|
| | `--quiet` (just print the Harvey–Collier test) |

Must follow the estimation of a model via OLS. Performs the CUSUM test—or if the `--squares` option is given, the CUSUMSQ test—for parameter stability. A series of one-step ahead forecast errors is obtained by running a series of regressions: the first regression uses the first $k$ observations and is used to generate a prediction of the dependent variable at observation $k + 1$; the

second uses the first $k + 1$ observations and generates a prediction for observation $k + 2$, and so on (where $k$ is the number of parameters in the original model).

The cumulated sum of the scaled forecast errors, or the squares of these errors, is printed and graphed. The null hypothesis of parameter stability is rejected at the 5 percent significance level if the cumulated sum strays outside of the 95 percent confidence band.

In the case of the CUSUM test, the Harvey–Collier $t$-statistic for testing the null hypothesis of parameter stability is also printed. See Greene's *Econometric Analysis* for details. For the CUSUMSQ test, the 95 percent confidence band is calculated using the algorithm given in Edgerton and Wells (1994).

Menu path: Model window, /Tests/CUSUM(SQ)


**data**

  Argument:    *varlist*
  Option:      `--quiet` (don't report results except on error)

Reads the variables in *varlist* from a database (gretl, RATS 4.0 or PcGive), which must have been opened previously using the open command. The data frequency and sample range may be established via the setobs and smpl commands prior to using this command. Here is a full example:

```
open macrodat.rat
setobs 4 1959:1
smpl ; 1999:4
data GDP_JP GDP_UK
```

The commands above open a database named `macrodat.rat`, establish a quarterly data set starting in the first quarter of 1959 and ending in the fourth quarter of 1999, and then import the series named GDP_JP and GDP_UK.

If `setobs` and `smpl` are not specified in this way, the data frequency and sample range are set using the first variable read from the database.

If the series to be read are of higher frequency than the working data set, you may specify a compaction method as below:

```
data (compact=average) LHUR PUNEW
```

The four available compaction methods are "average" (takes the mean of the high frequency observations), "last" (uses the last observation), "first" and "sum". If no method is specified, the default is to use the average.

Menu path: /File/Databases


**dataset**

  Arguments:    *keyword parameters*
  Examples:     `dataset addobs 24`
                 `dataset insobs 10`
                 `dataset compact 1`
                 `dataset compact 4 last`
                 `dataset expand interp`
                 `dataset transpose`
                 `dataset sortby x1`
                 `dataset resample 500`
                 `dataset renumber x 4`
                 `dataset clear`

Performs various operations on the data set as a whole, depending on the given *keyword*, which must be `addobs`, `insobs`, `clear`, `compact`, `expand`, `transpose`, `sortby`, `dsortby`, `resample` or `renumber`. Note: with the exception of `clear`, these actions are not available when the dataset is currently subsampled by selection of cases on some Boolean criterion.

`addobs`: Must be followed by a positive integer. Adds the specified number of extra observations to the end of the working dataset. This is primarily intended for forecasting purposes. The values of most variables over the additional range will be set to missing, but certain deterministic variables are recognized and extended, namely, a simple linear trend and periodic dummy variables.

`insobs`: Must be followed by a positive integer no greater than the current number of observations. Inserts a single observation at the specified position. All subsequent data are shifted by one place and the dataset is extended by one observation. All variables apart from the constant are given missing values at the new observation. This action is not available for panel datasets.

`clear`: No parameter required. Clears out the current data, returning gretl to its initial "empty" state.

`compact`: Must be followed by a positive integer representing a new data frequency, which should be lower than the current frequency (for example, a value of 4 when the current frequency is 12 indicates compaction from monthly to quarterly). This command is available for time series data only; it compacts all the series in the data set to the new frequency. A second parameter may be given, namely one of `sum`, `first` or `last`, to specify, respectively, compaction using the sum of the higher-frequency values, start-of-period values or end-of-period values. The default is to compact by averaging.

`expand`: This command is only available for annual or quarterly time series data: annual data can be expanded to quarterly, and quarterly data to monthly frequency. By default all the series in the data set are padded out to the new frequency by repeating the existing values, but if the modifier `interp` is appended then the series are expanded using Chow-Lin interpolation: the regressors are a constant and quadratic trend and an AR(1) disturbance process is assumed.

`transpose`: No additional parameter required. Transposes the current data set. That is, each observation (row) in the current data set will be treated as a variable (column), and each variable as an observation. This command may be useful if data have been read from some external source in which the rows of the data table represent variables.

`sortby`: The name of a single series or list is required. If one series is given, the observations on all variables in the dataset are re-ordered by increasing value of the specified series. If a list is given, the sort proceeds hierarchically: if the observations are tied in sort order with respect to the first key variable then the second key is used to break the tie, and so on until the tie is broken or the keys are exhausted. Note that this command is available only for undated data.

`dsortby`: Works as `sortby` except that the re-ordering is by decreasing value of the key series.

`resample`: Constructs a new dataset by random sampling, with replacement, of the rows of the current dataset. One argument is required, namely the number of rows to include. This may be less than, equal to, or greater than the number of observations in the original data. The original dataset can be retrieved via the command `smpl full`.

`renumber`: Requires the name of an existing series followed by an integer between 1 and the number of series in the dataset minus one. Moves the specified series to the specified position in the dataset, renumbering the other series accordingly. (Position 0 is occupied by the constant, which cannot be moved.)

Menu path: /Data

### debug

  Argument:   *function*

Experimental debugger for user-defined functions, available in the command-line program, gretlcli, and in the GUI console. The `debug` command should be invoked after the function in question is defined but before it is called. The effect is that execution pauses when the

function is called and a special prompt is shown.

At the debugging prompt you can type `next` to execute the next command in the function, or `continue` to allow execution of the function to continue unimpeded. These commands can be abbreviated as `n` and `c` respectively. You can also interpolate an instruction at this prompt, for example a `print` command to reveal the current value of some variable of interest.

### delete

| | |
|---|---|
| Argument: | *varname* |
| Option: | `--db` (delete series from database) |

This command is an all-purpose destructor for named variables (whether series, scalars, matrices, strings or bundles). It should be used with caution; no confirmation is asked.

In the case of series, *varname* may take the form of a named list, in which case all series in the list are deleted, or it may take the form of an explicit list of series given by name or ID number. Note that when you delete series any series with higher ID numbers than those on the deletion list will be re-numbered.

If the `--db` option is given, this command deletes the listed series not from the current dataset but from a gretl database, assuming that a database has been opened, and the user has write permission for file in question. See also the open command.

Menu path: Main window pop-up (single selection)

### diff

| | |
|---|---|
| Argument: | *varlist* |

The first difference of each variable in *varlist* is obtained and the result stored in a new variable with the prefix `d_`. Thus `diff x y` creates the new variables

```
d_x = x(t) - x(t-1)
d_y = y(t) - y(t-1)
```

Menu path: /Add/First differences of selected variables

### difftest

| | |
|---|---|
| Arguments: | *var1 var2* |
| Options: | `--sign` (Sign test, the default) |
| | `--rank-sum` (Wilcoxon rank-sum test) |
| | `--signed-rank` (Wilcoxon signed-rank test) |
| | `--verbose` (print extra output) |

Carries out a nonparametric test for a difference between two populations or groups, the specific test depending on the option selected.

With the `--sign` option, the Sign test is performed. This test is based on the fact that if two samples, $x$ and $y$, are drawn randomly from the same distribution, the probability that $x_i > y_i$, for each observation $i$, should equal 0.5. The test statistic is $w$, the number of observations for which $x_i > y_i$. Under the null hypothesis this follows the Binomial distribution with parameters $(n, 0.5)$, where $n$ is the number of observations.

With the `--rank-sum` option, the Wilcoxon rank-sum test is performed. This test proceeds by ranking the observations from both samples jointly, from smallest to largest, then finding the sum of the ranks of the observations from one of the samples. The two samples do not have to be of the same size, and if they differ the smaller sample is used in calculating the rank-sum. Under the null hypothesis that the samples are drawn from populations with the same median, the probability distribution of the rank-sum can be computed for any given sample sizes; and for reasonably large samples a close Normal approximation exists.

With the `--signed-rank` option, the Wilcoxon signed-rank test is performed. This is designed for matched data pairs such as, for example, the values of a variable for a sample of individuals before and after some treatment. The test proceeds by finding the differences between the paired observations, $x_i - y_i$, ranking these differences by absolute value, then assigning to each pair a signed rank, the sign agreeing with the sign of the difference. One then calculates $W_+$, the sum of the positive signed ranks. As with the rank-sum test, this statistic has a well-defined distribution under the null that the median difference is zero, which converges to the Normal for samples of reasonable size.

For the Wilcoxon tests, if the `--verbose` option is given then the ranking is printed. (This option has no effect if the Sign test is selected.)

### discrete

| | |
|---|---|
| Argument: | *varlist* |
| Option: | `--reverse` (mark variables as continuous) |

Marks each variable in *varlist* as being discrete. By default all variables are treated as continuous; marking a variable as discrete affects the way the variable is handled in frequency plots, and also allows you to select the variable for the command dummify.

If the `--reverse` flag is given, the operation is reversed; that is, the variables in *varlist* are marked as being continuous.

Menu path: /Variable/Edit attributes

### dpanel

| | |
|---|---|
| Argument: | *p* ; *depvar indepvars* [ ; *instruments* ] |
| Options: | `--quiet` (don't show estimated model) |
| | `--vcv` (print covariance matrix) |
| | `--two-step` (perform 2-step GMM estimation) |
| | `--system` (add equations in levels) |
| | `--time-dummies` (add time dummy variables) |
| | `--dpdstyle` (emulate DPD package for Ox) |
| | `--asymptotic` (uncorrected asymptotic standard errors) |
| Examples: | `dpanel 2 ; y x1 x2` |
| | `dpanel 2 ; y x1 x2 --system` |
| | `dpanel {2 3} ; y x1 x2 ; x1` |
| | `dpanel 1 ; y x1 x2 ; x1 GMM(x2,2,3)` |
| | See also `bbond98.inp` |

Carries out estimation of dynamic panel data models (that is, panel models including one or more lags of the dependent variable) using either the GMM-DIF or GMM-SYS method.

The parameter *p* represents the order of the autoregression for the dependent variable. In the simplest case this is a scalar value, but a pre-defined matrix may be given for this argument, to specify a set of (possibly non-contiguous) lags to be used.

The dependent variable and regressors should be given in levels form; they will be differenced automatically (since this estimator uses differencing to cancel out the individual effects).

The last (optional) field in the command is for specifying instruments. If no instruments are given, it is assumed that all the independent variables are strictly exogenous. If you specify any instruments, you should include in the list any strictly exogenous independent variables. For predetermined regressors, you can use the `GMM` function to include a specified range of lags in block-diagonal fashion. This is illustrated in the third example above. The first argument to `GMM` is the name of the variable in question, the second is the minimum lag to be used as an instrument, and the third is the maximum lag. The same syntax can be used with the `GMMlevel` function to specify GMM-type instruments for the equations in levels.

By default the results of 1-step estimation are reported (with robust standard errors). You may select 2-step estimation as an option. In both cases tests for autocorrelation of orders 1 and 2 are provided, as well as the Sargan overidentification test and a Wald test for the joint significance of the regressors. Note that in this differenced model first-order autocorrelation is not a threat to the validity of the model, but second-order autocorrelation violates the maintained statistical assumptions.

In the case of 2-step estimation, standard errors are by default computed using the finite-sample correction suggested by Windmeijer (2005). The standard asymptotic standard errors associated with the 2-step estimator are generally reckoned to be an unreliable guide to inference, but if for some reason you want to see them you can use the `--asymptotic` option to turn off the Windmeijer correction.

If the `--time-dummies` option is given, a set of time dummy variables is added to the specified regressors. The number of dummies is one less than the maximum number of periods used in estimation, to avoid perfect collinearity with the constant. The dummies are entered in differenced form unless the `--dpdstyle` option is given, in which case they are entered in levels.

For further details and examples, please see the *Gretl User's Guide*.

Menu path: /Model/Panel/Dynamic panel model


**dummify**

| | |
|---|---|
| Argument: | *varlist* |
| Options: | `--drop-first` (omit lowest value from encoding) |
| | `--drop-last` (omit highest value from encoding) |

For any suitable variables in *varlist*, creates a set of dummy variables coding for the distinct values of that variable. Suitable variables are those that have been explicitly marked as discrete, or those that take on a fairly small number of values all of which are "fairly round" (multiples of 0.25).

By default a dummy variable is added for each distinct value of the variable in question. For example if a discrete variable x has 5 distinct values, 5 dummy variables will be added to the data set, with names Dx_1, Dx_2 and so on. The first dummy variable will have value 1 for observations where x takes on its smallest value, 0 otherwise; the next dummy will have value 1 when x takes on its second-smallest value, and so on. If one of the option flags `--drop-first` or `--drop-last` is added, then either the lowest or the highest value of each variable is omitted from the encoding (which may be useful for avoiding the "dummy variable trap").

This command can also be embedded in the context of a regression specification. For example, the following line specifies a model where y is regressed on the set of dummy variables coding for x. (Option flags cannot be passed to dummify in this context.)

```
ols y dummify(x)
```


**duration**

| | |
|---|---|
| Arguments: | *depvar indepvars* [ ; *censvar* ] |
| Options: | `--exponential` (use exponential distribution) |
| | `--loglogistic` (use log-logistic distribution) |
| | `--lognormal` (use log-normal distribution) |
| | `--medians` (fitted values are medians) |
| | `--robust` (robust (QML) standard errors) |
| | `--vcv` (print covariance matrix) |
| | `--verbose` (print details of iterations) |
| Examples: | `duration y 0 x1 x2` |
| | `duration y 0 x1 x2 ; cens` |

Estimates a duration model: the dependent variable (which must be positive) represents the duration of some state of affairs, for example the length of spells of unemployment for a cross-section of respondents. By default the Weibull distribution is used but the exponential, log-logistic and log-normal distributions are also available.

If some of the duration measurements are right-censored (e.g. an individual's spell of unemployment has not come to an end within the period of observation) then you should supply the trailing argument *censvar*, a series in which non-zero values indicate right-censored cases.

By default the fitted values obtained via the accessor `$yhat` are the conditional means of the durations, but if the `--medians` option is given then `$yhat` provides the conditional medians instead.

Please see the *Gretl User's Guide* for details.

Menu path: /Model/Nonlinear models/Duration data...

### elif

See if.

### else

See if. Note that `else` requires a line to itself, before the following conditional command. You can append a comment, as in

```
else # OK, do something different
```

But you cannot append a command, as in

```
else x = 5 # wrong!
```

### end

Ends a block of commands of some sort. For example, `end system` terminates an equation system.

### endif

See if.

### endloop

Marks the end of a command loop. See loop.

### eqnprint

  Argument:    [ *-f filename* ]
  Option:      `--complete` (Create a complete document)

Must follow the estimation of a model. Prints the estimated model in the form of a LaTeX equation. If a filename is specified using the `-f` flag output goes to that file, otherwise it goes to a file with a name of the form `equation_N.tex`, where `N` is the number of models estimated to date in the current session. See also tabprint.

If the `--complete` flag is given, the LaTeX file is a complete document, ready for processing; otherwise it must be included in a document.

Menu path: Model window, /LaTeX

**equation**

| Arguments: | *depvar indepvars* |
|---|---|
| Example: | `equation y x1 x2 x3 const` |

Specifies an equation within a system of equations (see system). The syntax for specifying an equation within an SUR system is the same as that for, e.g., ols. For an equation within a Three-Stage Least Squares system you may either (a) give an OLS-type equation specification and provide a common list of instruments using the `instr` keyword (again, see system), or (b) use the same equation syntax as for tsls.

**estimate**

| Arguments: | [ *systemname* ] [ *estimator* ] |
|---|---|
| Options: | `--iterate` (iterate to convergence) |
| | `--no-df-corr` (no degrees of freedom correction) |
| | `--geomean` (see below) |
| | `--quiet` (don't print results) |
| | `--verbose` (print details of iterations) |
| Examples: | `estimate "Klein Model 1" method=fiml` |
| | `estimate Sys1 method=sur` |
| | `estimate Sys1 method=sur --iterate` |

Calls for estimation of a system of equations, which must have been previously defined using the system command. The name of the system should be given first, surrounded by double quotes if the name contains spaces. The estimator, which must be one of ols, tsls, sur, 3sls, fiml or liml, is preceded by the string `method=`. These arguments are optional if the system in question has already been estimated and occupies the place of the "last model"; in that case the estimator defaults to the previously used value.

If the system in question has had a set of restrictions applied (see the restrict command), estimation will be subject to the specified restrictions.

If the estimation method is sur or 3sls and the `--iterate` flag is given, the estimator will be iterated. In the case of SUR, if the procedure converges the results are maximum likelihood estimates. Iteration of three-stage least squares, however, does not in general converge on the full-information maximum likelihood results. The `--iterate` flag is ignored for other methods of estimation.

If the equation-by-equation estimators ols or tsls are chosen, the default is to apply a degrees of freedom correction when calculating standard errors. This can be suppressed using the `--no-df-corr` flag. This flag has no effect with the other estimators; no degrees of freedom correction is applied in any case.

By default, the formula used in calculating the elements of the cross-equation covariance matrix is

$$\hat{\sigma}_{i,j} = \frac{\hat{u}_i' \hat{u}_j}{T}$$

If the `--geomean` flag is given, a degrees of freedom correction is applied: the formula is

$$\hat{\sigma}_{i,j} = \frac{\hat{u}_i' \hat{u}_j}{\sqrt{(T - k_i)(T - k_j)}}$$

where the $k$s denote the number of independent parameters in each equation.

If the `--verbose` option is given and an iterative method is specified, details of the iterations are printed.

**fcast**

| | |
|---|---|
| Arguments: | [ *startobs endobs* ] [ *steps-ahead* ] [ *varname* ] |
| Options: | `--dynamic` (create dynamic forecast) |
| | `--static` (create static forecast) |
| | `--out-of-sample` (generate post-sample forecast) |
| | `--no-stats` (don't print forecast statistics) |
| | `--quiet` (don't print anything) |
| | `--rolling` (see below) |
| | `--plot`[=*filename*] (see below) |
| Examples: | `fcast 1997:1 2001:4 f1` |
| | `fcast fit2` |
| | `fcast 2004:1 2008:3 4 rfcast --rolling` |

Must follow an estimation command. Forecasts are generated for a certain range of observations: if *startobs* and *endobs* are given, for that range (if possible); otherwise if the `--out-of-sample` option is given, for observations following the range over which the model was estimated; otherwise over the currently defined sample range. If an out-of-sample forecast is requested but no relevant observations are available, an error is flagged. Depending on the nature of the model, standard errors may also be generated; see below. Also see below for the special effect of the `--rolling` option.

If the last model estimated is a single equation, then the optional *varname* argument has the following effect: the forecast values are not printed, but are saved to the dataset under the given name. If the last model is a system of equations, *varname* has a different effect, namely selecting a particular endogenous variable for forecasting (the default being to produce forecasts for all the endogenous variables). In the system case, or if *varname* is not given, the forecast values can be retrieved using the accessor `$fcast`, and the standard errors, if available, via `$fcerr`.

The choice between a static and a dynamic forecast applies only in the case of dynamic models, with an autoregressive error process or including one or more lagged values of the dependent variable as regressors. Static forecasts are one step ahead, based on realized values from the previous period, while dynamic forecasts employ the chain rule of forecasting. For example, if a forecast for $y$ in 2008 requires as input a value of $y$ for 2007, a static forecast is impossible without actual data for 2007. A dynamic forecast for 2008 is possible if a prior forecast can be substituted for $y$ in 2007.

The default is to give a static forecast for any portion of the forecast range that lies within the sample range over which the model was estimated, and a dynamic forecast (if relevant) out of sample. The `--dynamic` option requests a dynamic forecast from the earliest possible date, and the `--static` option requests a static forecast even out of sample.

The `--rolling` option is presently available only for single-equation models estimated via OLS. When this option is given the forecasts are recursive. That is, each forecast is generated from an estimate of the given model using data from a fixed starting point (namely, the start of the sample range for the original estimation) up to the forecast date minus $k$, where $k$ is the number of steps ahead, which must be given in the *steps-ahead* argument. The forecasts are always dynamic if this is applicable. Note that the *steps-ahead* argument should be given only in conjunction with the `--rolling` option.

The `--plot` option (available only in the case of single-equation estimation) calls for a plot file to be produced, containing a graphical representation of the forecast. When no *filename* parameter is given, gretl writes a gnuplot command file with a name on the pattern `gpttmp01.plt` to the user's gretl working directory (with the number incremented for successive plots). If a *filename* is appended, its extension is used to determine the type of file to be written (`.eps` for EPS, `.pdf` for PDF, or `.png` for PNG; any other extension gives a gnuplot command file). For example,

```
fcast --plot=fc.pdf
```

will generate a graphic in PDF format. Absolute pathnames are respected, otherwise files are

written to the gretl working directory.

The nature of the forecast standard errors (if available) depends on the nature of the model and the forecast. For static linear models standard errors are computed using the method outlined by Davidson and MacKinnon (2004); they incorporate both uncertainty due to the error process and parameter uncertainty (summarized in the covariance matrix of the parameter estimates). For dynamic models, forecast standard errors are computed only in the case of a dynamic forecast, and they do not incorporate parameter uncertainty. For nonlinear models, forecast standard errors are not presently available.

Menu path: Model window, /Analysis/Forecasts

### foreign

| Syntax: | `foreign language=`*lang* |
|---------|---------|
| Options: | `--send-data` (pre-load the current dataset; see below) |
| | `--quiet` (suppress output from foreign program) |

This command opens a special mode in which commands to be executed by another program are accepted. You exit this mode with `end foreign`; at this point the stacked commands are executed.

At present three "foreign" programs are supported in this way, GNU R (`language=R`), Jurgen Doornik's Ox (`language=Ox`) and GNU Octave (`language=Octave`). Language names are recognized on a case-insensitive basis.

The `--send-data` option is valid only in connection with R and Octave; it has the effect of making the entire current gretl dataset available within the target program, under the name `gretldata`.

See the *Gretl User's Guide* for details and examples.

### fractint

| Arguments: | *series* [ *order* ] |
|------------|---------|
| Options: | `--gph` (do Geweke and Porter-Hudak test) |
| | `--all` (do both tests) |
| | `--quiet` (don't print results) |

Tests the specified series for fractional integration ("long memory"). The null hypothesis is that the integration order of the series is zero. By default the local Whittle estimator (Robinson, 1995) is used but if the `--gph` option is given the GPH test (Geweke and Porter-Hudak, 1983) is performed instead. If the `--all` flag is given then the results of both tests are printed.

For details on this sort of test, see Phillips and Shimotsu (2004).

If the optional *order* argument is not given the order for the test(s) is set automatically as the lesser of $T/2$ and $T^{0.6}$.

The results can be retrieved using the accessors `$test` and `$pvalue`. These values are based on the Local Whittle Estimator unless the `--gph` option is given.

Menu path: /Variable/Unit root tests/Fractional integration

**freq**

| | |
|---|---|
| Argument: | *series* |
| Options: | `--nbins=`*n* (specify number of bins) |
| | `--min=`*minval* (specify minimum, see below) |
| | `--binwidth=`*width* (specify bin width, see below) |
| | `--quiet` (suppress printing of graph) |
| | `--normal` (test for the normal distribution) |
| | `--gamma` (test for gamma distribution) |
| | `--silent` (don't print anything) |
| | `--show-plot` (see below) |
| Examples: | `freq x` |
| | `freq x --normal` |
| | `freq x --nbins=5` |
| | `freq x --min=0 --binwidth=0.10` |

With no options given, displays the frequency distribution for *series* (given by name or number), with the number of bins and their size chosen automatically.

To control the presentation of the distribution you may specify *either* the number of bins or the minimum value plus the width of the bins, as shown in the last two examples above. The `--min` option sets the lower limit of the left-most bin.

If the `--normal` option is given, the Doornik–Hansen chi-square test for normality is computed. If the `--gamma` option is given, the test for normality is replaced by Locke's nonparametric test for the null hypothesis that the variable follows the gamma distribution; see Locke (1976), Shapiro and Chen (2001). Note that the parameterization of the gamma distribution used in gretl is (shape, scale).

In interactive mode a graph of the distribution is displayed by default. The `--quiet` flag can be used to suppress this. Conversely, the graph is not usually shown when the `freq` is used in a script, but you can force its display by giving the `--show-plot` option. (This does not apply when using the command-line program, `gretlcli`.)

The `--silent` flag suppresses the usual output entirely. This makes sense only in conjunction with one or other of the distribution test options: the test statistic and its p-value are recorded, and can be retrieved using the accessors `$test` and `$pvalue`.

Menu path: /Variable/Frequency distribution

**function**

| | |
|---|---|
| Argument: | *fnname* |

Opens a block of statements in which a function is defined. This block must be closed with `end function`. Please see the *Gretl User's Guide* for details.

**garch**

| | |
|---|---|
| Arguments: | *p q* ; *depvar* [ *indepvars* ] |
| Options: | `--robust` (robust standard errors) |
| | `--verbose` (print details of iterations) |
| | `--vcv` (print covariance matrix) |
| | `--nc` (do not include a constant) |
| | `--stdresid` (standardize the residuals) |
| | `--fcp` (use Fiorentini, Calzolari, Panattoni algorithm) |
| | `--arma-init` (initial variance parameters from ARMA) |
| Examples: | `garch 1 1 ; y` |
| | `garch 1 1 ; y 0 x1 x2 --robust` |

Estimates a GARCH model (GARCH = Generalized Autoregressive Conditional Heteroskedasticity), either a univariate model or, if *indepvars* are specified, including the given exogenous variables. The integer values $p$ and $q$ (which may be given in numerical form or as the names of pre-existing scalar variables) represent the lag orders in the conditional variance equation:

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}$$

The parameter $p$ therefore represents the Generalized (or "AR") order, while $q$ represents the regular ARCH (or "MA") order. If $p$ is non-zero, $q$ must also be non-zero otherwise the model is unidentified. However, you can estimate a regular ARCH model by setting $q$ to a positive value and $p$ to zero. The sum of $p$ and $q$ must be no greater than 5. Note that a constant is automatically included in the mean equation unless the `--nc` option is given.

By default native gretl code is used in estimation of GARCH models, but you also have the option of using the algorithm of Fiorentini *et al.* (1996). The former uses the BFGS maximizer while the latter uses the information matrix to maximize the likelihood, with fine-tuning via the Hessian.

Several variant estimators of the covariance matrix are available with this command. By default, the Hessian is used unless the `--robust` option is given, in which case the QML (White) covariance matrix is used. Other possibilities (e.g. the information matrix, or the Bollerslev–Wooldridge estimator) can be specified using the set command.

By default, the estimates of the variance parameters are initialized using the unconditional error variance from initial OLS estimation for the constant, and small positive values for the coefficients on the past values of the squared error and the error variance. The flag `--arma-init` calls for the starting values of these parameters to be set using an initial ARMA model, exploiting the relationship between GARCH and ARMA set out in Chapter 21 of Hamilton's *Time Series Analysis*. In some cases this may improve the chances of convergence.

The GARCH residuals and estimated conditional variance can be retrieved as `$uhat` and `$h` respectively. For example, to get the conditional variance:

```
series ht = $h
```

If the `--stdresid` option is given, the `$uhat` values are divided by the square root of $h_t$.

Menu path: /Model/Time series/GARCH

**genr**

   Arguments:    *newvar = formula*

In the appropriate context, `series`, `scalar` and `matrix` are synonyms for this command.

Creates new variables, often via transformations of existing variables. See also diff, logs, lags, ldiff, sdiff and square for shortcuts. In the context of a `genr` formula, existing variables must be referenced by name, not ID number. The formula should be a well-formed combination of variable names, constants, operators and functions (described below). Note that further details on some aspects of this command can be found in the *Gretl User's Guide*.

```
series c = 10
```

A `genr` command may yield either a series or a scalar result. For example, the formula `x2 = x * 2` naturally yields a series if the variable `x` is a series and a scalar if `x` is a scalar. The formulae `x = 0` and `mx = mean(x)` naturally return scalars. Under some circumstances you may want to have a scalar result expanded into a series or vector. You can do this by using `series` as an "alias" for the `genr` command. For example, `series x = 0` produces a series all of whose values are set to 0. You can also use `scalar` as an alias for `genr`. It is not possible to coerce a vector result into a scalar, but use of this keyword indicates that the result *should be* a scalar: if it is not, an error occurs.

When a formula yields a series result, the range over which the result is written to the target variable depends on the current sample setting. It is possible, therefore, to define a series piecewise using the `smpl` command in conjunction with `genr`.

Supported *arithmetical operators* are, in order of precedence: ^ (exponentiation); *, / and % (modulus or remainder); + and -.

The available *Boolean operators* are (again, in order of precedence): ! (negation), && (logical AND), || (logical OR), >, <, =, >= (greater than or equal), <= (less than or equal) and != (not equal). The Boolean operators can be used in constructing dummy variables: for instance (x > 10) returns 1 if x > 10, 0 otherwise.

Built-in constants are `pi` and `NA`. The latter is the missing value code: you can initialize a variable to the missing value with `scalar x = NA`.

The `genr` command supports a wide range of mathematical and statistical functions, including all the common ones plus several that are special to econometrics. In addition it offers access to numerous internal variables that are defined in the course of running regressions, doing hypothesis tests, and so on.

For a listing of functions and accessors, see Chapter 2.

Besides the operators and functions noted above there are some special uses of `genr`:

- `genr time` creates a time trend variable (1,2,3,...) called `time`. `genr index` does the same thing except that the variable is called `index`.

- `genr dummy` creates dummy variables up to the periodicity of the data. In the case of quarterly data (periodicity 4), the program creates `dq1` = 1 for first quarter and 0 in other quarters, `dq2` = 1 for the second quarter and 0 in other quarters, and so on. With monthly data the dummies are named `dm1`, `dm2`, and so on. With other frequencies the names are `dummy_1`, `dummy_2`, etc.

- `genr unitdum` and `genr timedum` create sets of special dummy variables for use with panel data. The first codes for the cross-sectional units and the second for the time period of the observations.

*Note*: In the command-line program, `genr` commands that retrieve model-related data always reference the model that was estimated most recently. This is also true in the GUI program, if one uses `genr` in the "gretl console" or enters a formula using the "Define new variable" option under the Variable menu in the main window. With the GUI, however, you have the option of retrieving data from any model currently displayed in a window (whether or not it's the most recent model). You do this under the "Model data" menu in the model's window.

The special variable `obs` serves as an index of the observations. For instance `genr dum = (obs=15)` will generate a dummy variable that has value 1 for observation 15, 0 otherwise. You can also use this variable to pick out particular observations by date or name. For example, `genr d = (obs>1986:4)`, `genr d = (obs>"2008/04/01")`, or `genr d = (obs="CA")`. If daily dates or observation labels are used in this context, they should be enclosed in double quotes. Quarterly and monthly dates (with a colon) may be used unquoted. Note that in the case of annual time series data, the year is not distinguishable syntactically from a plain integer; therefore if you wish to compare observations against `obs` by year you must use the function `obsnum` to convert the year to a 1-based index value, as in `genr d = (obs>obsnum(1986))`.

Scalar values can be pulled from a series in the context of a `genr` formula, using the syntax *varname*[*obs*]. The *obs* value can be given by number or date. Examples: x[5], CPI[1996:01]. For daily data, the form *YYYY/MM/DD* should be used, e.g. ibm[1970/01/23].

An individual observation in a series can be modified via `genr`. To do this, a valid observation number or date, in square brackets, must be appended to the name of the variable on the left-hand side of the formula. For example, `genr x[3] = 30` or `genr x[1950:04] = 303.7`.

Menu path: /Add/Define new variable

Other access: Main window pop-up menu

**Table 1.1**: Examples of use of genr command

| *Formula* | *Comment* |
|---|---|
| `y = x1^3` | x1 cubed |
| `y = ln((x1+x2)/x3)` | |
| `z = x>y` | $z(t) = 1$ if `x(t) > y(t)`, otherwise 0 |
| `y = x(-2)` | x lagged 2 periods |
| `y = x(+2)` | x led 2 periods |
| `y = diff(x)` | `y(t) = x(t) - x(t-1)` |
| `y = ldiff(x)` | `y(t) = log x(t) - log x(t-1)`, the instantaneous rate of growth of x |
| `y = sort(x)` | sorts x in increasing order and stores in y |
| `y = dsort(x)` | sort x in decreasing order |
| `y = int(x)` | truncate x and store its integer value as y |
| `y = abs(x)` | store the absolute values of x |
| `y = sum(x)` | sum x values excluding missing NA entries |
| `y = cum(x)` | cumulation: $y_t = \sum_{\tau=1}^{t} x_\tau$ |
| `aa = $ess` | set `aa` equal to the Error Sum of Squares from last regression |
| `x = $coeff(sqft)` | grab the estimated coefficient on the variable `sqft` from the last regression |
| `rho4 = $rho(4)` | grab the 4th-order autoregressive coefficient from the last model (presumes an `ar` model) |
| `cvx1x2 = $vcv(x1, x2)` | grab the estimated coefficient covariance of vars `x1` and `x2` from the last model |
| `foo = uniform()` | uniform pseudo-random variable in range 0–1 |
| `bar = 3 * normal()` | normal pseudo-random variable, $\mu = 0$, $\sigma = 3$ |
| `samp = ok(x)` | = 1 for observations where x is not missing. |

**gmm**

Options:  `--two-step` (two step estimation)

`--iterate` (iterated GMM)

`--vcv` (print covariance matrix)

`--verbose` (print details of iterations)

`--lbfgs` (use L-BFGS-B instead of regular BFGS)

Performs Generalized Method of Moments (GMM) estimation using the BFGS (Broyden, Fletcher, Goldfarb, Shanno) algorithm. You must specify one or more commands for updating the relevant quantities (typically GMM residuals), one or more sets of orthogonality conditions, an initial matrix of weights, and a listing of the parameters to be estimated, all enclosed between the tags `gmm` and `end gmm`. Any options should be appended to the `end gmm` line.

Please see the *Gretl User's Guide* for details on this command. Here we just illustrate with a simple example.

```
gmm e = y - X*b
  orthog e ; W
  weights V
  params b
end gmm
```

In the example above we assume that `y` and `X` are data matrices, `b` is an appropriately sized vector of parameter values, `W` is a matrix of instruments, and `V` is a suitable matrix of weights. The statement

```
orthog e ; W
```

indicates that the residual vector `e` is in principle orthogonal to each of the instruments composing the columns of `W`.

Menu path: /Model/GMM


**gnuplot**

| | |
|---|---|
| Arguments: | *yvars xvar* [ *dumvar* ] |
| Options: | `--with-lines`[=*varspec*] (use lines, not points) |
| | `--with-impulses`[=*varspec*] (use vertical lines) |
| | `--time-series` (plot against time) |
| | `--suppress-fitted` (don't show fitted line) |
| | `--single-yaxis` (force use of just one y-axis) |
| | `--linear-fit` (show least squares fit) |
| | `--inverse-fit` (show inverse fit) |
| | `--quadratic-fit` (show quadratic fit) |
| | `--cubic-fit` (show cubic fit) |
| | `--loess-fit` (show loess fit) |
| | `--semilog-fit` (show semilog fit) |
| | `--dummy` (see below) |
| | `--matrix=`*name* (plot columns of named matrix) |
| | `--output=`*filename* (send output to specified file) |
| | `--input=`*filename* (take input from specified file) |
| Examples: | `gnuplot y1 y2 x` |
| | `gnuplot x --time-series --with-lines` |
| | `gnuplot wages educ gender --dummy` |
| | `gnuplot y1 y2 x --with-lines=y2` |

The variables in the list *yvars* are graphed against *xvar*. For a time series plot you may either give `time` as *xvar* or use the option flag `--time-series`.

By default, data-points are shown as points; this can be overridden by giving one of the options `--with-lines` or `--with-impulses`. If more than one variable is to be plotted on the $y$ axis, the effect of these options may be confined to a subset of the variables by using the *varspec* parameter. This should take the form of a comma-separated listing of the names or numbers of the variables to be plotted with lines or impulses respectively. For instance, the final example above shows how to plot `y1` and `y2` against `x`, such that `y2` is represented by a line but `y1` by points.

If the `--dummy` option is selected, exactly three variables should be given: a single $y$ variable, an $x$ variable, and *dvar*, a discrete variable. The effect is to plot *yvar* against *xvar* with the points shown in different colors depending on the value of *dvar* at the given observation.

Generally, the arguments *yvars* and *xvar* are required, and refer to series in the current dataset (given either by name or ID number). But if a named matrix is supplied via the `--matrix` option these arguments become optional: if the specified matrix has $k$ columns, by default the first $k$ − 1 columns are treated as the *yvars* and the last column as *xvar*. If the `--time-series` option is given, however, all $k$ columns are plotted against time. If you wish to plot selected columns of the matrix, you should specify *yvars* and *xvar* in the form of 1-based column numbers. For example if you want a scatterplot of column 2 of matrix `M` against column 1, you can do:

```
gnuplot 2 1 --matrix=M
```

In interactive mode the plot is displayed immediately. In batch mode the default behavior is that a gnuplot command file is written in the user's working directory, with a name on the pattern `gpttmpN.plt`, starting with N = 01. The actual plots may be generated later using gnuplot (under MS Windows, wgnuplot). This behavior can be modified by use of the `--output=`*filename*

option. This option controls the filename used, and at the same time allows you to specify a particular output format via the three-letter extension of the file name, as follows: `.eps` results in the production of an Encapsulated PostScript (EPS) file; `.pdf` produces PDF; `.png` produces PNG format, `.emf` calls for EMF (Enhanced MetaFile), `.fig` calls for an Xfig file, and `.svg` for SVG (Scalable Vector Graphics). If the dummy filename "`display`" is given then the plot is shown on screen as in interactive mode. If a filename with any extension other than those just mentioned is given, a gnuplot command file is written.

The various "fit" options are applicable only for bivariate scatterplots and single time-series plots. The default behavior for a scatterplot is to show the OLS fit if the slope coefficient is significant at the 10 percent level; this can be suppressed via the `--suppress-fitted` option. The default behavior for time-series is not to show a fitted line. If the `--linear` option is given, the OLS line is shown regardless of whether or not it is significant. The other fit options (`--inverse`, `--quadratic`, `--cubic`, `--loess` and `--semilog`) produce respectively an inverse fit (regression of $y$ on $1/x$), a quadratic fit, a cubic fit, a loess fit, and a semilog fit. Loess (also sometimes called "lowess") is a robust locally weighted regression. By semilog, we mean a regression of log $y$ on $x$ (or time); the fitted line represents the conditional expectation of $y$, obtained by exponentiation.

A further option to this command is available: following the specification of the variables to be plotted and the option flag (if any), you may add literal gnuplot commands to control the appearance of the plot (for example, setting the plot title and/or the axis ranges). These commands should be enclosed in braces, and each gnuplot command must be terminated with a semi-colon. A backslash may be used to continue a set of gnuplot commands over more than one line. Here is an example of the syntax:

```
{ set title 'My Title'; set yrange [0:1000]; }
```

Menu path: /View/Graph specified vars

Other access: Main window pop-up menu, graph button on toolbar

### graphpg

The session "graph page" will work only if you have the LaTeX typesetting system installed, and are able to generate and view PDF or PostScript output.

In the session icon window, you can drag up to eight graphs onto the graph page icon. When you double-click on the graph page (or right-click and select "Display"), a page containing the selected graphs will be composed and opened in a suitable viewer. From there you should be able to print the page.

To clear the graph page, right-click on its icon and select "Clear".

In script (or console) mode, you can add a graph to the graph page by issuing the command `graphpg add` after saving a named graph, as in

```
grf1 <- gnuplot Y X
graphpg add
```

Also in script mode you can call for display of the graph page using the command `graphpg show`, and can clear the page via `graphpg free`. In addition you can adjust the scale of the font used in the graph page via `graphpg fontscale` *scale*, where *scale* is a multiplier (with a default of 1.0). Thus to make the font size 50 percent bigger than the default you can do

```
graphpg fontscale 1.5
```

Note that on systems other than MS Windows, you may have to adjust the setting for the program used to view PDF or PostScript files. Find that under the "Programs" tab in the gretl Preferences dialog box (under the Tools menu in the main window).

**hausman**

This test is available only after estimating an OLS model using panel data (see also `setobs`). It tests the simple pooled model against the principal alternatives, the fixed effects and random effects models.

The fixed effects model allows the intercept of the regression to vary across the cross-sectional units. An *F*-test is reported for the null hypotheses that the intercepts do not differ. The random effects model decomposes the residual variance into two parts, one part specific to the cross-sectional unit and the other specific to the particular observation. (This estimator can be computed only if the number of cross-sectional units in the data set exceeds the number of parameters to be estimated.) The Breusch–Pagan LM statistic tests the null hypothesis that the pooled OLS estimator is adequate against the random effects alternative.

The pooled OLS model may be rejected against both of the alternatives, fixed effects and random effects. Provided the unit- or group-specific error is uncorrelated with the independent variables, the random effects estimator is more efficient than the fixed effects estimator; otherwise the random effects estimator is inconsistent and the fixed effects estimator is to be preferred. The null hypothesis for the Hausman test is that the group-specific error is not so correlated (and therefore the random effects model is preferable). A low p-value for this test counts against the random effects model and in favor of fixed effects.

Menu path: Model window, /Tests/Panel diagnostics

**heckit**

| | |
|---|---|
| Arguments: | *depvar indepvars* ; *selection equation* |
| Options: | `--quiet` (suppress printing of results) |
| | `--robust` (QML standard errors) |
| | `--two-step` (perform two-step estimation) |
| | `--vcv` (print covariance matrix) |
| | `--verbose` (print extra output) |
| Example: | `heckit y 0 x1 x2 ; ys 0 x3 x4` |
| | `heckit.inp` |

Heckman-type selection model. In the specification, the list before the semicolon represents the outcome equation, and the second list represents the selection equation. The dependent variable in the selection equation (`ys` in the example above) must be a binary variable.

By default, the parameters are estimated by maximum likelihood. The covariance matrix of the parameters is computed using the negative inverse of the Hessian. If two-step estimation is desired, use the `--two-step` option. In this case, the covariance matrix of the parameters of the outcome equation is appropriately adjusted as per Heckman (1979).

Please note that in ML estimation a numerical approximation of the Hessian is used; this may lead to inaccuracies in the estimated covariance matrix if the scale of the explanatory variables is such that some of the estimated coefficients are very small in absolute value. This problem will be addressed in future versions; in the meantime, rescaling the offending explanatory variable(s) can be used as a workaround.

Menu path: /Model/Nonlinear models/Heckit

**help**

| | |
|---|---|
| Variants: | `help` |
| | `help functions` |
| | `help` *command* |
| | `help` *function* |
| Option: | `--func` (select functions help) |

If no arguments are given, prints a list of available commands. If the single argument `functions` is given, prints a list of available functions (see genr).

`help` *command* describes *command* (e.g. `help smpl`). `help` *function* describes *function* (e.g. `help ldet`). Some functions have the same names as related commands (e.g. `diff`): in that case the default is to print help for the command, but you can get help on the function by using the `--func` option.

Menu path: /Help

### hsk

| | |
|---|---|
| Arguments: | *depvar indepvars* |
| Option: | `--vcv` (print covariance matrix) |

This command is applicable where heteroskedasticity is present in the form of an unknown function of the regressors which can be approximated by a quadratic relationship. In that context it offers the possibility of consistent standard errors and more efficient parameter estimates as compared with OLS.

The procedure involves (a) OLS estimation of the model of interest, followed by (b) an auxiliary regression to generate an estimate of the error variance, then finally (c) weighted least squares, using as weight the reciprocal of the estimated variance.

In the auxiliary regression (b) we regress the log of the squared residuals from the first OLS on the original regressors and their squares. The log transformation is performed to ensure that the estimated variances are non-negative. Call the fitted values from this regression $u^*$. The weight series for the final WLS is then formed as $1/\exp(u^*)$.

Menu path: /Model/Other linear models/Heteroskedasticity corrected

### hurst

| | |
|---|---|
| Argument: | *series* |

Calculates the Hurst exponent (a measure of persistence or long memory) for a time-series variable having at least 128 observations.

The Hurst exponent is discussed by Mandelbrot. In theoretical terms it is the exponent, $H$, in the relationship

$$\mathrm{RS}(x) = an^H$$

where RS is the "rescaled range" of the variable $x$ in samples of size $n$ and $a$ is a constant. The rescaled range is the range (maximum minus minimum) of the cumulated value or partial sum of $x$ over the sample period (after subtraction of the sample mean), divided by the sample standard deviation.

As a reference point, if $x$ is white noise (zero mean, zero persistence) then the range of its cumulated "wandering" (which forms a random walk), scaled by the standard deviation, grows as the square root of the sample size, giving an expected Hurst exponent of 0.5. Values of the exponent significantly in excess of 0.5 indicate persistence, and values less than 0.5 indicate anti-persistence (negative autocorrelation). In principle the exponent is bounded by 0 and 1, although in finite samples it is possible to get an estimated exponent greater than 1.

In gretl, the exponent is estimated using binary sub-sampling: we start with the entire data range, then the two halves of the range, then the four quarters, and so on. For sample sizes smaller than the data range, the RS value is the mean across the available samples. The exponent is then estimated as the slope coefficient in a regression of the log of RS on the log of sample size.

Menu path: /Variable/Hurst exponent

### if

Flow control for command execution. Three sorts of construction are supported, as follows.

```
# simple form
if condition
    commands
endif
```

```
# two branches
if condition
    commands1
else
    commands2
endif
```

```
# three or more branches
if condition1
    commands1
elif condition2
    commands2
else
    commands3
endif
```

*condition* must be a Boolean expression, for the syntax of which see genr. More than one `elif` block may be included. In addition, `if ... endif` blocks may be nested.

### include

| | |
|---|---|
| Argument: | *filename* |
| Examples: | `include myfile.inp` |
| | `include sols.gfn` |

Intended for use in a command script, primarily for including definitions of functions. Executes the commands in *filename* then returns control to the main script. To include a packaged function, be sure to include the filename extension.

See also run.

### info

Prints out any supplementary information stored with the current datafile.

Menu path: /Data/Dataset info

Other access: Data browser windows

### intreg

| | |
|---|---|
| Arguments: | *minvar maxvar indepvars* |
| Options: | `--quiet` (suppress printing of results) |
| | `--verbose` (print details of iterations) |
| | `--robust` (robust standard errors) |
| Example: | `intreg lo hi const x1 x2` |
| | `wtp.inp` |

Estimates an interval regression model. This model arises when the dependent variable is imperfectly observed for some (possibly all) observations. In other words, the data generating process is assumed to be

$$y_t^* = x_t \beta + \epsilon_t$$

but we only observe

$$m_t \leq y_t \leq M_t$$

(the interval may be left- or right-unbounded). Note that for some observations $m$ may equal $M$. The variables *minvar* and *maxvar* must contain NAs for left- and right-unbounded observations, respectively.

The model is estimated by maximum likelihood, assuming normality of the disturbance term.

By default, standard errors are computed using the negative inverse of the Hessian. If the `--robust` flag is given, then QML or Huber–White standard errors are calculated instead. In this case the estimated covariance matrix is a "sandwich" of the inverse of the estimated Hessian and the outer product of the gradient.

Menu path: /Model/Nonlinear models/Interval regression

**kalman**

Options:    `--cross` (allow for cross-correlated disturbances)
            `--diffuse` (use diffuse initialization)

Opens a block of statements to set up a Kalman filter. This block should end with the line `end kalman`, to which the options shown above may be appended. The intervening lines specify the matrices that compose the filter. For example,

```
kalman
  obsy y
  obsymat H
  statemat F
  statevar Q
end kalman
```

Please see the *Gretl User's Guide* for details.

See also kfilter, ksimul, ksmooth.

**kpss**

Arguments:  *order varlist*
Options:    `--trend` (include a trend)
            `--seasonals` (include seasonal dummies)
            `--verbose` (print regression results)
            `--quiet` (suppress printing of results)
            `--difference` (use first difference of variable)
Examples:   `kpss 8 y`
            `kpss 4 x1 --trend`

For use of this command with panel data please see the final section in this entry.

Computes the KPSS test (Kwiatkowski *et al.*, 1992) for stationarity, for each of the specified variables (or their first difference, if the `--difference` option is selected). The null hypothesis is that the variable in question is stationary, either around a level or, if the `--trend` option is given, around a deterministic linear trend.

The order argument determines the size of the window used for Bartlett smoothing. If the `--verbose` option is chosen the results of the auxiliary regression are printed, along with the estimated variance of the random walk component of the variable.

The critical values shown for the test statistic are based on the response surfaces estimated by Sephton (1995), which are more accurate for small samples than the values given in the original KPSS article. When the test statistic lies between the 10 percent and 1 percent critical values a p-value is shown; this is obtained by linear interpolation and should not be taken too literally.

*Panel data*

When the kpss command is used with panel data, to produce a panel unit root test, the applicable options and the results shown are somewhat different. While you may give a list of variables for testing in the regular time-series case, with panel data only one variable may be tested per command. And the --verbose option has a different meaning: it produces a brief account of the test for each individual time series (the default being to show only the overall result).

When possible, the overall test (null hypothesis: the series in question is stationary for all the panel units) is calculated using the method of Choi (2001). This is not always straightforward, the difficulty being that while the Choi test is based on the p-values of the tests on the individual series, we do not currently have a means of calculating p-values for the KPSS test statistic; we must rely on a few critical values.

If the test statistic for a given series falls between the 10 percent and 1 percent critical values, we are able to interpolate a p-value. But if the test falls short of the 10 percent value, or exceeds the 1 percent value, we cannot interpolate and can at best place a bound on the global Choi test. If the individual test statistic falls short of the 10 percent value for some units but exceeds the 1 percent value for others, we cannot even compute a bound for the global test.

Menu path: /Variable/Unit root tests/KPSS test

## labels

|            |                               |
|------------|-------------------------------|
| Variants:  | labels [ *varlist* ]          |
|            | labels --to-file=*filename*   |
|            | labels --from-file=*filename* |

In the first form, prints out the informative labels (if present) for the series in *varlist*, or for all series in the dataset if *varlist* is not specified.

With the option --to-file, writes to the named file the labels for all series in the dataset, one per line. If no labels are present an error is flagged; if some series have labels and others do not, a blank line is printed for series with no label.

With the option --from-file, reads the specified file (which should be plain text) and assigns labels to the series in the dataset, reading one label per line and taking blank lines to indicate blank labels.

Via the file writing and reading options it is possible to copy series labels from one data file to another.

## lad

|            |                               |
|------------|-------------------------------|
| Arguments: | *depvar indepvars*            |
| Option:    | --vcv (print covariance matrix) |

Calculates a regression that minimizes the sum of the absolute deviations of the observed from the fitted values of the dependent variable. Coefficient estimates are derived using the Barrodale–Roberts simplex algorithm; a warning is printed if the solution is not unique.

Standard errors are derived using the bootstrap procedure with 500 drawings. The covariance matrix for the parameter estimates, printed when the --vcv flag is given, is based on the same bootstrap.

Menu path: /Model/Robust estimation/Least Absolute Deviation

## lags

|            |                          |
|------------|--------------------------|
| Variants:  | lags *varlist*           |
|            | lags *order* ; *varlist* |
| Examples:  | lags x y                 |
|            | lags 12 ; x y            |

Creates new series which are lagged values of each of the series in *varlist*. By default the number of lags created equals the periodicity of the data. For example, if the periodicity is 4 (quarterly), the command `lags x` creates

```
x_1 = x(t-1)
x_2 = x(t-2)
x_3 = x(t-3)
x_4 = x(t-4)
```

The number of lags created can be controlled by the optional first parameter.

Menu path: /Add/Lags of selected variables

### ldiff

  Argument:   *varlist*

The first difference of the natural log of each series in *varlist* is obtained and the result stored in a new series with the prefix `ld_`. Thus `ldiff x y` creates the new variables

```
ld_x = log(x) - log(x(-1))
ld_y = log(y) - log(y(-1))
```

Menu path: /Add/Log differences of selected variables

### leverage

  Options:   `--save` (save variables)
             `--quiet` (don't print results)

Must follow an `ols` command. Calculates the leverage ($h$, which must lie in the range 0 to 1) for each data point in the sample on which the previous model was estimated. Displays the residual ($u$) for each observation along with its leverage and a measure of its influence on the estimates, $uh/(1-h)$. "Leverage points" for which the value of $h$ exceeds $2k/n$ (where $k$ is the number of parameters being estimated and $n$ is the sample size) are flagged with an asterisk. For details on the concepts of leverage and influence see Davidson and MacKinnon (1993), Chapter 2.

DFFITS values are also computed: these are "studentized residuals" (predicted residuals divided by their standard errors) multiplied by $\sqrt{h/(1-h)}$. For discussions of studentized residuals and DFFITS see chapter 12 of Maddala (1992) or Belsley *et al.* (1980).

Briefly, a "predicted residual" is the difference between the observed value of the dependent variable at observation $t$, and the fitted value for observation $t$ obtained from a regression in which that observation is omitted (or a dummy variable with value 1 for observation $t$ alone has been added); the studentized residual is obtained by dividing the predicted residual by its standard error.

If the `--save` flag is given with this command, then the leverage, influence and DFFITS values are added to the current data set. In that context the `--quiet` flag may be used to suppress the printing of results.

After execution, the `$test` accessor returns the cross-validation criterion, which is defined as

$$\sum_{i=1}^{n}(y_i - \hat{y}_{-i})^2$$

where $\hat{y}_{-i}$ is the forecast error for the $i$-th observation, after it has been excluded from the sample. The criterion is, hence, the sum of the squared forecasting errors where all $n$ observations but the $i$-th one are used to predict it (the so-called *leave-one-out* estimator). For a broader discussion of the cross-validation criterion, see Davidson and MacKinnon's *Econometric Theory and Methods*, pages 685–686, and the references therein.

Menu path: Model window, /Tests/Influential observations

**levinlin**

| | |
|---|---|
| Arguments: | *order series* |
| Options: | `--nc` (test without a constant) |
| | `--ct` (with constant and trend) |
| | `--quiet` (suppress printing of results) |
| Examples: | `levinlin 0 y` |
| | `levinlin 2 y --ct` |
| | `levinlin {2,2,3,3,4,4} y` |

Carries out the panel unit-root test described by Levin *et al.* (2002). The null hypothesis is that all of the individual time series exhibit a unit root, and the alternative is that none of the series has a unit root. (That is, a common AR(1) coefficient is assumed, although in other respects the statistical properties of the series are allowed to vary across individuals.)

By default the test ADF regressions include a constant; to suppress the constant use the `--nc` option, or to add a linear trend use the `--ct` option. (See the adf command for explanation of ADF regressions.)

The (non-negative) *order* for the test (governing the number of lags of the dependent variable to include in the ADF regressions) may be given in either of two forms. If a scalar value is given, this is applied to all the individuals in the panel. The alternative is to provide a matrix containing a specific lag order for each individual; this must be a vector with as many elements as there are individuals in the current sample range. Such a matrix can be specified by name, or constructed using braces as illustrated in the last example above.

Menu path: /Variable/Unit root tests/Levin-Lin-Chu test

**logistic**

| | |
|---|---|
| Arguments: | *depvar indepvars* |
| Options: | `--ymax=`*value* (specify maximum of dependent variable) |
| | `--vcv` (print covariance matrix) |
| Examples: | `logistic y const x` |
| | `logistic y const x --ymax=50` |

Logistic regression: carries out an OLS regression using the logistic transformation of the dependent variable,

$$\log\left(\frac{y}{y^* - y}\right)$$

The dependent variable must be strictly positive. If all its values lie between 0 and 1, the default is to use a $y^*$ value (the asymptotic maximum of the dependent variable) of 1; if its values lie between 0 and 100, the default $y^*$ is 100.

If you wish to set a different maximum, use the `--ymax` option. Note that the supplied value must be greater than all of the observed values of the dependent variable.

The fitted values and residuals from the regression are automatically transformed using

$$y = \frac{y^*}{1 + e^{-x}}$$

where $x$ represents either a fitted value or a residual from the OLS regression using the transformed dependent variable. The reported values are therefore comparable with the original dependent variable.

Note that if the dependent variable is binary, you should use the logit command instead.

Menu path: /Model/Nonlinear models/Logistic

**logit**

Arguments:   *depvar indepvars*
Options:     `--robust` (robust standard errors)
             `--multinomial` (estimate multinomial logit)
             `--vcv` (print covariance matrix)
             `--verbose` (print details of iterations)
             `--p-values` (show p-values instead of slopes)

If the dependent variable is a binary variable (all values are 0 or 1) maximum likelihood estimates of the coefficients on *indepvars* are obtained via the Newton–Raphson method. As the model is nonlinear the slopes depend on the values of the independent variables. By default the slopes with respect to each of the independent variables are calculated (at the means of those variables) and these slopes replace the usual p-values in the regression output. This behavior can be suppressed my giving the `--p-values` option. The chi-square statistic tests the null hypothesis that all coefficients are zero apart from the constant.

By default, standard errors are computed using the negative inverse of the Hessian. If the `--robust` flag is given, then QML or Huber–White standard errors are calculated instead. In this case the estimated covariance matrix is a "sandwich" of the inverse of the estimated Hessian and the outer product of the gradient. See chapter 10 of Davidson and MacKinnon for details.

If the dependent variable is not binary but is discrete, then by default it is interpreted as an ordinal response, and Ordered Logit estimates are obtained. However, if the `--multinomial` option is given, the dependent variable is interpreted as an unordered response, and Multinomial Logit estimates are produced. (In either case, if the variable selected as dependent is not discrete an error is flagged.) In the multinomial case, the accessor `$mnlprobs` is available after estimation, to get a matrix containing the estimated probabilities of the outcomes at each observation (observations in rows, outcomes in columns).

If you want to use logit for analysis of proportions (where the dependent variable is the proportion of cases having a certain characteristic, at each observation, rather than a 1 or 0 variable indicating whether the characteristic is present or not) you should not use the `logit` command, but rather construct the logit variable, as in

```
series lgt_p = log(p/(1 - p))
```

and use this as the dependent variable in an OLS regression. See chapter 12 of Ramanathan (2002).

Menu path: /Model/Nonlinear models/Logit

**logs**

Argument:   *varlist*

The natural log of each of the series in *varlist* is obtained and the result stored in a new series with the prefix `l_` ("el" underscore). For example, `logs x y` creates the new variables `l_x` = ln(x) and `l_y` = ln(y).

Menu path: /Add/Logs of selected variables

**loop**

| | |
|---|---|
| Argument: | *control* |
| Options: | --progressive (enable special forms of certain commands) |
| | --verbose (report details of genr commands) |
| | --quiet (do not report number of iterations performed) |
| Examples: | loop 1000 |
| | loop 1000 --progressive |
| | loop while essdiff > .00001 |
| | loop i=1991..2000 |
| | loop for (r=-.99; r<=.99; r+=.01) |
| | loop foreach i xlist |

This command opens a special mode in which the program accepts commands to be executed repeatedly. You exit the mode of entering loop commands with endloop: at this point the stacked commands are executed.

The parameter *control* may take any of five forms, as shown in the examples: an integer number of times to repeat the commands within the loop; "while" plus a boolean condition; a range of integer values for index variable; "for" plus three expressions in parentheses, separated by semicolons (which emulates the for statement in the C programming language); or "foreach" plus an index variable and a list.

See the *Gretl User's Guide* for further details and examples. The effect of the --progressive option (which is designed for use in Monte Carlo simulations) is explained there. Not all gretl commands may be used within a loop; the commands available in this context are also set out there.

**mahal**

| | |
|---|---|
| Argument: | *varlist* |
| Options: | --quiet (don't print anything) |
| | --save (add distances to the dataset) |
| | --vcv (print covariance matrix) |

The Mahalanobis distance is the distance between two points in a $k$-dimensional space, scaled by the statistical variation in each dimension of the space. For example, if $p$ and $q$ are two observations on a set of $k$ variables with covariance matrix $C$, then the Mahalanobis distance between the observations is given by

$$\sqrt{(p-q)'C^{-1}(p-q)}$$

where $(p-q)$ is a $k$-vector. This reduces to Euclidean distance if the covariance matrix is the identity matrix.

The space for which distances are computed is defined by the selected variables. For each observation in the current sample range, the distance is computed between the observation and the centroid of the selected variables. This distance is the multidimensional counterpart of a standard $z$-score, and can be used to judge whether a given observation "belongs" with a group of other observations.

If the --vcv option is given, the covariance matrix and its inverse are printed. If the --save option is given, the distances are saved to the dataset under the name mdist (or mdist1, mdist2 and so on if there is already a variable of that name).

Menu path: /View/Mahalanobis distances

### makepkg

| | |
|---|---|
| Argument: | *filename* |
| Options: | `--index` (write auxiliary index file) |
| | `--translations` (write auxiliary strings file) |

Supports creation of a gretl function package via the command line. The filename argument represents the name of the package to be created, and should have the `.gfn` extension. Please see the *Gretl User's Guide* for details.

The option flags support the writing of auxiliary files for use with gretl "addons". The index file is a short XML document containing basic information about the package; it has the same basename as the package and the extension `.xml`. The translations file contains strings from the package that may be suitable for translation, in C format; for package `foo` this file is named `foo-i18n.c`.

Menu path: /File/Function files/New package

### meantest

| | |
|---|---|
| Arguments: | *var1 var2* |
| Option: | `--unequal-vars` (assume variances are unequal) |

Calculates the $t$ statistic for the null hypothesis that the population means are equal for the variables *var1* and *var2*, and shows its p-value.

By default the test statistic is calculated on the assumption that the variances are equal for the two variables; with the `--unequal-vars` option the variances are assumed to be different. This will make a difference to the test statistic only if there are different numbers of non-missing observations for the two variables.

Menu path: /Model/Bivariate tests/Difference of means

### mle

| | |
|---|---|
| Arguments: | *log-likelihood function* [ *derivatives* ] |
| Options: | `--quiet` (don't show estimated model) |
| | `--vcv` (print covariance matrix) |
| | `--hessian` (base covariance matrix on the Hessian) |
| | `--robust` (QML covariance matrix) |
| | `--verbose` (print details of iterations) |
| | `--no-gradient-check` (see below) |
| | `--lbfgs` (use L-BFGS-B instead of regular BFGS) |
| Example: | `weibull.inp` |

Performs Maximum Likelihood (ML) estimation using either the BFGS (Broyden, Fletcher, Goldfarb, Shanno) algorithm or Newton's method. The user must specify the log-likelihood function. The parameters of this function must be declared and given starting values (using the `genr` command) prior to estimation. Optionally, the user may specify the derivatives of the log-likelihood function with respect to each of the parameters; if analytical derivatives are not supplied, a numerical approximation is computed.

Simple example: Suppose we have a series X with values 0 or 1 and we wish to obtain the maximum likelihood estimate of the probability, p, that X = 1. (In this simple case we can guess in advance that the ML estimate of p will simply equal the proportion of Xs equal to 1 in the sample.)

The parameter p must first be added to the dataset and given an initial value. For example, `scalar p = 0.5`.

We then construct the MLE command block:

```
mle loglik = X*log(p) + (1-X)*log(1-p)
   deriv p = X/p - (1-X)/(1-p)
end mle
```

The first line above specifies the log-likelihood function. It starts with the keyword `mle`, then a dependent variable is specified and an expression for the log-likelihood is given (using the same syntax as in the `genr` command). The next line (which is optional) starts with the keyword `deriv` and supplies the derivative of the log-likelihood function with respect to the parameter p. If no derivatives are given, you should include a statement using the keyword `params` which identifies the free parameters: these are listed on one line, separated by spaces and can be either scalars, or vectors, or any combination of the two. For example, the above could be changed to:

```
mle loglik = X*log(p) + (1-X)*log(1-p)
   params p
end mle
```

in which case numerical derivatives would be used.

Note that any option flags should be appended to the ending line of the MLE block.

By default, estimated standard errors are based on the Outer Product of the Gradient. If the `--hessian` option is given, they are instead based on the negative inverse of the Hessian (which is approximated numerically). If the `--robust` option is given, a QML estimator is used (namely, a sandwich of the negative inverse of the Hessian and the covariance matrix of the gradient).

If you supply analytical derivatives, by default gretl runs a numerical check on their plausibility. Occasionally this may produce false positives, instances where correct derivatives appear to be wrong and estimation is refused. To counter this, or to achieve a little extra speed, you can give the option `--no-gradient-check`. Obviously, you should do this only if you are quite confident that the gradient you have specified is right.

For a much more in-depth description of `mle`, please refer to the *Gretl User's Guide*.

Menu path: /Model/Maximum likelihood

### modeltab

  Arguments:   *add* or *show* or *free*

Manipulates the gretl "model table". See the *Gretl User's Guide* for details. The sub-commands have the following effects: `add` adds the last model estimated to the model table, if possible; `show` displays the model table in a window; and `free` clears the table.

Menu path: Session window, Model table icon

### modprint

  Arguments:   *coeffmat names* [ *addstats* ]

Prints the coefficient table and optional additional statistics for a model estimated "by hand". Mainly useful for user-written functions.

The argument *coeffmat* should be a $k$ by 2 matrix containing $k$ coefficients and $k$ associated standard errors, and *names* should be a string containing at least $k$ names for the coefficients, separated by commas or spaces. (The *names* argument may be either the name of a string variable or a literal string, enclosed in double quotes.)

The optional argument *addstats* is a vector containing $p$ additional statistics to be printed under the coefficient table. If this argument is given, then *names* should contain $k + p$ comma-separated strings, the additional $p$ strings to be associated with the additional statistics.

**modtest**

  Argument:    [ *order* ]

  Options:     --normality (normality of residual)

               --logs (non-linearity, logs)

               --autocorr (serial correlation)

               --arch (ARCH)

               --squares (non-linearity, squares)

               --white (heteroskedasticity, White's test)

               --white-nocross (White's test, squares only)

               --breusch-pagan (heteroskedasticity, Breusch–Pagan)

               --robust (robust variance estimate for Breusch–Pagan)

               --panel (heteroskedasticity, groupwise)

               --comfac (common factor restriction, AR1 models only)

               --quiet (don't print details)

Must immediately follow an estimation command. Depending on the option given, this command carries out one of the following: the Doornik–Hansen test for the normality of the error term; a Lagrange Multiplier test for nonlinearity (logs or squares); White's test (with or without cross-products) or the Breusch–Pagan test for heteroskedasticity; the LMF test for serial correlation (Kiviet, 1986); a test for ARCH (Autoregressive Conditional Heteroskedasticity; see also the arch command); or a test of the common factor restriction implied by AR(1) estimation. With the exception of the normality and common factor test most of the options are only available for models estimated via OLS, but see below for details regarding two-stage least squares.

The optional order argument is relevant only in case the --autocorr or --arch options are selected. The default is to run these tests using a lag order equal to the periodicity of the data, but this can be adjusted by supplying a specific lag order.

The --robust option applies only when the Breusch–Pagan test is selected; its effect is to use the robust variance estimator proposed by Koenker (1981), making the test less sensitive to the assumption of normality.

The --panel option is available only when the model is estimated on panel data: in this case a test for groupwise heteroskedasticity is performed (that is, for a differing error variance across the cross-sectional units).

The --comfac option is available only when the model is estimated via an AR(1) method such as Hildreth–Lu. The auxiliary regression takes the form of a relatively unrestricted dynamic model, which is used to test the common factor restriction implicit in the AR(1) specification.

By default, the program prints the auxiliary regression on which the test statistic is based, where applicable. This may be suppressed by using the --quiet flag. The test statistic and its p-value may be retrieved using the accessors $test and $pvalue respectively.

When a model has been estimated by two-stage least squares (see tsls), the LM principle breaks down and gretl offers some equivalents: the --autocorr option computes Godfrey's test for autocorrelation (Godfrey, 1994) while the --white option yields the HET1 heteroskedasticity test (Pesaran and Taylor, 1999).

Menu path: Model window, /Tests

**mpols**

  Arguments:   *depvar indepvars*

  Options:     --vcv (print covariance matrix)

              --simple-print (do not print auxiliary statistics)

              --quiet (suppress printing of results)

Computes OLS estimates for the specified model using multiple precision floating-point arithmetic. This command is available only if gretl is compiled with support for the Gnu Multiple

Precision (GMP) library. By default 256 bits of precision are used for the calculations, but this can be increased via the environment variable GRETL_MP_BITS. For example, when using the bash shell one could issue the following command, before starting gretl, to set a precision of 1024 bits.

```
export GRETL_MP_BITS=1024
```

A rather arcane option is available for this command (primarily for testing purposes): if the *indepvars* list is followed by a semicolon and a further list of numbers, those numbers are taken as powers of $x$ to be added to the regression, where $x$ is the last variable in *indepvars*. These additional terms are computed and stored in multiple precision. In the following example y is regressed on x and the second, third and fourth powers of x:

```
mpols y 0 x ; 2 3 4
```

Menu path: /Model/Other linear models/High precision OLS

**negbin**

| | |
|---|---|
| Arguments: | *depvar indepvars* [ ; *offset* ] |
| Options: | --model1 (use NegBin 1 model) |
| | --opg (see below) |
| | --robust (QML covariance matrix) |
| | --vcv (print covariance matrix) |
| | --verbose (print details of iterations) |

Estimates a Negative Binomial model. The dependent variable is taken to represent a count of the occurrence of events of some sort, and must have only non-negative integer values. By default the model NegBin 2 is used, in which the conditional variance of the count is given by $\mu(1 + \alpha\mu)$, where $\mu$ denotes the conditional mean. But if the --model1 option is given the conditional variance is $\mu(1 + \alpha)$.

The optional offset series works in the same way as for the poisson command. The Poisson model is a restricted form of the Negative Binomial in which $\alpha = 0$ by construction.

By default, standard errors are computed using a numerical approximation to the Hessian at convergence. But if the --opg option is given the covariance matrix is based on the Outer Product of the Gradient (OPG), or if the --robust option is given QML standard errors are calculated, using a "sandwich" of the inverse of the Hessian and the OPG.

Menu path: /Model/Nonlinear models/Count data...

**nls**

| | |
|---|---|
| Arguments: | *function* [ *derivatives* ] |
| Options: | --quiet (don't show estimated model) |
| | --robust (robust standard errors) |
| | --vcv (print covariance matrix) |
| | --verbose (print details of iterations) |
| Example: | wg_nls.inp |

Performs Nonlinear Least Squares (NLS) estimation using a modified version of the Levenberg–Marquardt algorithm. You must supply a function specification. The parameters of this function must be declared and given starting values (using the genr command) prior to estimation. Optionally, you may specify the derivatives of the regression function with respect to each of the parameters. If you do not supply derivatives you should instead give a list of the parameters to be estimated (separated by spaces or commas), preceded by the keyword params. In the latter case a numerical approximation to the Jacobian is computed.

It is easiest to show what is required by example. The following is a complete script to estimate the nonlinear consumption function set out in William Greene's *Econometric Analysis* (Chapter 11 of the 4th edition, or Chapter 9 of the 5th). The numbers to the left of the lines are for reference and are not part of the commands. Note that any option flags, such as `--vcv` for printing the covariance matrix of the parameter estimates, should be appended to the final command, `end nls`.

```
1   open greene11_3.gdt
2   ols C 0 Y
3   scalar a = $coeff(0)
4   scalar b = $coeff(Y)
5   scalar g = 1.0
6   nls C = a + b * Y^g
7     deriv a = 1
8     deriv b = Y^g
9     deriv g = b * Y^g * log(Y)
10  end nls --vcv
```

It is often convenient to initialize the parameters by reference to a related linear model; that is accomplished here on lines 2 to 5. The parameters alpha, beta and gamma could be set to any initial values (not necessarily based on a model estimated with OLS), although convergence of the NLS procedure is not guaranteed for an arbitrary starting point.

The actual NLS commands occupy lines 6 to 10. On line 6 the `nls` command is given: a dependent variable is specified, followed by an equals sign, followed by a function specification. The syntax for the expression on the right is the same as that for the `genr` command. The next three lines specify the derivatives of the regression function with respect to each of the parameters in turn. Each line begins with the keyword `deriv`, gives the name of a parameter, an equals sign, and an expression whereby the derivative can be calculated (again, the syntax here is the same as for `genr`). As an alternative to supplying numerical derivatives, you could substitute the following for lines 7 to 9:

```
params a b g
```

Line 10, `end nls`, completes the command and calls for estimation. Any options should be appended to this line.

For further details on NLS estimation please see the *Gretl User's Guide*.

Menu path: /Model/Nonlinear models/Nonlinear Least Squares

**normtest**

| | |
|---|---|
| Argument: | *series* |
| Options: | `--dhansen` (Doornik–Hansen test, the default) |
| | `--swilk` (Shapiro–Wilk test) |
| | `--lillie` (Lilliefors test) |
| | `--jbera` (Jarque–Bera test) |
| | `--all` (do all tests) |
| | `--quiet` (suppress printed output) |

Carries out a test for normality for the given *series*. The specific test is controlled by the option flags (but if no flag is given, the Doornik-Hansen test is performed). Note: the Doornik-Hansen and Shapiro–Wilk tests are recommended over the others, on account of their superior small-sample properties.

The test statistic and its p-value may be retrieved using the accessors `$test` and `$pvalue`. Please note that if the `--all` option is given, the result recorded is that from the Doornik–Hansen test.

Menu path: /Variable/Normality test

**nulldata**

| | |
|---|---|
| Argument: | *series-length* |
| Option: | `--preserve` (preserve matrices) |
| Example: | `nulldata 500` |

Establishes a "blank" data set, containing only a constant and an index variable, with periodicity 1 and the specified number of observations. This may be used for simulation purposes: some of the `genr` commands (e.g. `genr uniform()`, `genr normal()`) will generate dummy data from scratch to fill out the data set. This command may be useful in conjunction with `loop`. See also the "seed" option to the set command.

By default, this command cleans out all data in gretl's current workspace.  If you give the `--preserve` option, however, any currently defined matrices are retained.

Menu path: /File/New data set

**ols**

| | |
|---|---|
| Arguments: | *depvar indepvars* |
| Options: | `--vcv` (print covariance matrix) |
| | `--robust` (robust standard errors) |
| | `--jackknife` (see below) |
| | `--simple-print` (do not print auxiliary statistics) |
| | `--quiet` (suppress printing of results) |
| | `--anova` (print an ANOVA table) |
| | `--no-df-corr` (suppress degrees of freedom correction) |
| | `--print-final` (see below) |
| Examples: | `ols 1 0 2 4 6 7` |
| | `ols y 0 x1 x2 x3 --vcv` |
| | `ols y 0 x1 x2 x3 --quiet` |

Computes ordinary least squares (OLS) estimates with *depvar* as the dependent variable and *indepvars* as the list of independent variables. Variables may be specified by name or number; use the number zero for a constant term.

Besides coefficient estimates and standard errors, the program also prints p-values for $t$ (two-tailed) and $F$-statistics.  A p-value below 0.01 indicates statistical significance at the 1 percent level and is marked with ***. ** indicates significance between 1 and 5 percent and * indicates significance between the 5 and 10 percent levels.  Model selection statistics (the Akaike Information Criterion or AIC and Schwarz's Bayesian Information Criterion) are also printed. The formula used for the AIC is that given by Akaike (1974), namely minus two times the maximized log-likelihood plus two times the number of parameters estimated.

If the option `--no-df-corr` is given, the usual degrees of freedom correction is not applied when calculating the estimated error variance (and hence also the standard errors of the parameter estimates).

The option `--print-final` is applicable only in the context of a loop.  It arranges for the regression to be run silently on all but the final iteration of the loop. See the *Gretl User's Guide* for details.

Various internal variables may be retrieved following estimation. For example

```
series uh = $uhat
```

saves the residuals under the name `uh`. See the "accessors" section of the gretl function reference for details.

The specific formula used for generating robust standard errors (when the `--robust` option is given) can be adjusted via the set command. The `--jackknife` option has the effect of selecting an `hc_version` of 3a; it is provided to emulate the old `hccm` command.

Menu path: /Model/Ordinary Least Squares

Other access: Beta-hat button on toolbar

### omit

| | |
|---|---|
| Argument: | *varlist* |
| Options: | `--test-only` (don't replace the current model) |
| | `--chi-square` (give chi-square form of Wald test) |
| | `--quiet` (print only the basic test result) |
| | `--silent` (don't print anything) |
| | `--vcv` (print covariance matrix for reduced model) |
| | `--auto[=`*alpha*`]` (sequential elimination, see below) |
| Examples: | `omit 5 7 9` |
| | `omit seasonals --quiet` |
| | `omit --auto` |
| | `omit --auto=0.05` |

This command must follow an estimation command. It calculates a Wald test for the joint significance of the variables in *varlist*, which should be a subset of the independent variables in the model last estimated. The results of the test may be retrieved using the accessors `$test` and `$pvalue`.

By default the restricted model is estimated and it replaces the original as the "current model" for the purposes of, for example, retrieving the residuals as `$uhat` or doing further tests. This behavior may be suppressed via the `--test-only` option.

By default the *F*-form of the Wald test is recorded; the `--chi-square` option may be used to record the chi-square form instead.

If the restricted model is both estimated and printed, the `--vcv` option has the effect of printing its covariance matrix, otherwise this option is ignored.

Alternatively, if the `--auto` flag is given, sequential elimination is performed: at each step the variable with the highest p-value is omitted, until all remaining variables have a p-value no greater than some cutoff. The default cutoff is 10 percent (two-sided); this can be adjusted by appending "=" and a value between 0 and 1 (with no spaces), as in the fourth example above. If *varlist* is given this process is confined to the listed variables, otherwise all variables are treated as candidates for omission. Note that the `--auto` and `--test-only` options cannot be combined.

Menu path: Model window, /Tests/Omit variables

### open

| | |
|---|---|
| Argument: | *filename* |
| Options: | `--quiet` (don't print list of series) |
| | `--preserve` (preserve any matrices and scalars) |
| | `--www` (use a database on the gretl server) |
| | See below for additional specialized options |
| Examples: | `open data4-1` |
| | `open voter.dta` |
| | `open fedbog --www` |

Opens a data file. If a data file is already open, it is replaced by the newly opened one. To add data to the current dataset, see append.

If a full path is not given, the program will search some relevant paths to try to find the file. If no filename suffix is given (as in the first example above), gretl assumes a native datafile with

suffix `.gdt`. Based on the name of the file and various heuristics, gretl will try to detect the format of the data file (native, plain text, CSV, MS Excel, Stata, SPSS, etc.).

If the *filename* argument takes the form of a URI starting with `http://`, then gretl will attempt to download the indicated data file before opening it.

By default, opening a new data file clears the current gretl session, which includes deletion of any named matrices and scalars. If you wish to keep any currently defined matrices and scalars, use the `--preserve` option.

The `open` command can also be used to open a database (gretl, RATS 4.0 or PcGive) for reading. In that case it should be followed by the data command to extract particular series from the database. If the `www` option is given, the program will try to access a database of the given name on the gretl server — for instance the Federal Reserve interest rates database in the third example above.

When opening a spreadsheet file (Gnumeric, Open Document or MS Excel), you may give up to three additional parameters following the filename. First, you can select a particular worksheet within the file. This is done either by giving its (1-based) number, using the syntax, e.g., `--sheet=2`, or, if you know the name of the sheet, by giving the name in double quotes, as in `--sheet="MacroData"`. The default is to read the first worksheet. You can also specify a column and/or row offset into the worksheet via, e.g.,

        --coloffset=3 --rowoffset=2

which would cause gretl to ignore the first 3 columns and the first 2 rows. The default is an offset of 0 in both dimensions, that is, to start reading at the top-left cell.

With plain text files, gretl generally expects to find the data columns delimited in some standard manner. But there is also a special facility for reading "fixed format" files, in which there are no delimiters but there is a known specification of the form, e.g., "variable $k$ occupies 8 columns starting at column 24". To read such files, you should append a string `--cols=`*colspec*, where *colspec* is composed of comma-separated integers. These integers are interpreted as a set of pairs. The first element of each pair denotes a starting column, measured in bytes from the beginning of the line with 1 indicating the first byte; and the second element indicates how many bytes should be read for the given field. So, for example, if you say

        open fixed.txt --cols=1,6,20,3

then for variable 1 gretl will read 6 bytes starting at column 1; and for variable 2, 3 bytes starting at column 20. Lines that are blank, or that begin with `#`, are ignored, but otherwise the column-reading template is applied, and if anything other than a valid numerical value is found an error is flagged. If the data are read successfully, the variables will be named `v1`, `v2`, etc. It's up to the user to provide meaningful names and/or descriptions using the commands rename and/or setinfo.

Menu path: /File/Open data

Other access: Drag a data file into gretl (MS Windows or Gnome)


**orthdev**

  Argument:    *varlist*

Applicable with panel data only. A series of forward orthogonal deviations is obtained for each variable in *varlist* and stored in a new variable with the prefix `o_`. Thus `orthdev x y` creates the new variables `o_x` and `o_y`.

The values are stored one step ahead of their true temporal location (that is, `o_x` at observation $t$ holds the deviation that, strictly speaking, belongs at $t − 1$). This is for compatibility with first differences: one loses the first observation in each time series, not the last.

**outfile**

| | |
|---|---|
| Variants: | `outfile` *filename option* |
| | `outfile --close` |
| Options: | `--append` (append to file) |
| | `--write` (overwrite file) |
| Examples: | `outfile regress.txt --write` |
| | `outfile --close` |

Diverts output to *filename*, until further notice. Use the flag `--append` to append output to an existing file or `--write` to start a new file (or overwrite an existing one). Only one file can be opened in this way at any given time.

The `--close` flag is used to close an output file that was previously opened as above. Output will then revert to the default stream.

In the first example command above, the file `regress.txt` is opened for writing, and in the second it is closed. This would make sense as a sequence only if some commands were issued before the `--close`. For example if an `ols` command intervened, its output would go to `regress.txt` rather than the screen.

Three special variants on the above are available. If you give the keyword `null` in place of a real filename along with the `--write` option, the effect is to suppress all printed output until the next `outfile --close`. In addition if the keywords `stdout` or `stderr` are given in place of a regular filename the effect is to redirect output to standard output or standard error output respectively.

**panel**

| | |
|---|---|
| Arguments: | *depvar indepvars* |
| Options: | `--vcv` (print covariance matrix) |
| | `--fixed-effects` (estimate with group fixed effects) |
| | `--random-effects` (random effects or GLS model) |
| | `--nerlove` (use the Nerlove transformation) |
| | `--between` (estimate the between-groups model) |
| | `--robust` (robust standard errors; see below) |
| | `--time-dummies` (include time dummy variables) |
| | `--unit-weights` (weighted least squares) |
| | `--iterate` (iterative estimation) |
| | `--matrix-diff` (use matrix-difference method for Hausman test) |
| | `--quiet` (less verbose output) |
| | `--verbose` (more verbose output) |

Estimates a panel model. By default the fixed effects estimator is used; this is implemented by subtracting the group or unit means from the original data.

If the `--random-effects` flag is given, random effects estimates are computed, by default using the method of Swamy and Arora (1972). In this case (only) the option `--matrix-diff` forces use of the matrix-difference method (as opposed to the regression method) for carrying out the Hausman test for the consistency of the random effects estimator. Also specific to the random effects estimator is the `--nerlove` flag, which selects the method of Nerlove (1971) as opposed to Swamy and Arora.

Alternatively, if the `--unit-weights` flag is given, the model is estimated via weighted least squares, with the weights based on the residual variance for the respective cross-sectional units in the sample. In this case (only) the `--iterate` flag may be added to produce iterative estimates: if the iteration converges, the resulting estimates are Maximum Likelihood.

As a further alternative, if the `--between` flag is given, the between-groups model is estimated (that is, an OLS regression using the group means).

The `--robust` option is available only for fixed effects models. The default variant is the Arellano HAC estimator, but Beck–Katz "Panel Corrected Standard Errors" can be selected via the command `set pcse on`.

For more details on panel estimation, please see the *Gretl User's Guide*.

Menu path: /Model/Panel

**pca**

| | |
|---|---|
| Argument: | *varlist* |
| Options: | `--covariance` (use the covariance matrix) |
| | `--save[=`*n*`]` (save major components) |
| | `--save-all` (save all components) |
| | `--quiet` (don't print results) |

Principal Components Analysis. Unless the `--quiet` option is given, prints the eigenvalues of the correlation matrix (or the covariance matrix if the `--covariance` option is given) for the variables in *varlist*, along with the proportion of the joint variance accounted for by each component. Also prints the corresponding eigenvectors (or "component loadings").

If you give the `--save-all` option then all components are saved to the dataset as series, with names PC1, PC2 and so on. These artificial variables are formed as the sum of (component loading) times (standardized $X_i$), where $X_i$ denotes the $i$th variable in *varlist*.

If you give the `--save` option without a parameter value, components with eigenvalues greater than the mean (which means greater than 1.0 if the analysis is based on the correlation matrix) are saved to the dataset as described above. If you provide a value for $n$ with this option then the most important $n$ components are saved.

See also the princomp function.

Menu path: /View/Principal components

Other access: Main window pop-up (multiple selection)

**pergm**

| | |
|---|---|
| Arguments: | *series* [ *bandwidth* ] |
| Options: | `--bartlett` (use Bartlett lag window) |
| | `--log` (use log scale) |
| | `--radians` (show frequency in radians) |
| | `--degrees` (show frequency in degrees) |

Computes and displays (and if not in batch mode, graphs) the spectrum of the specified series. By default the sample periodogram is given, but optionally a Bartlett lag window is used in estimating the spectrum (see, for example, Greene's *Econometric Analysis* for a discussion of this). The default width of the Bartlett window is twice the square root of the sample size but this can be set manually using the *bandwidth* parameter, up to a maximum of half the sample size.

If the `--log` option is given the spectrum is represented on a logarithmic scale.

The (mutually exclusive) options `--radians` and `--degrees` influence the appearance of the frequency axis when the periodogram is graphed. By default the frequency is scaled by the number of periods in the sample, but these options cause the axis to be labeled from 0 to $\pi$ radians or from 0 to 180°, respectively.

Menu path: /Variable/Periodogram

Other access: Main window pop-up menu (single selection)

**poisson**

| | |
|---|---|
| Arguments: | *depvar indepvars* [ ; *offset* ] |
| Options: | `--robust` (robust standard errors) |
| | `--vcv` (print covariance matrix) |
| | `--verbose` (print details of iterations) |
| Examples: | `poisson y 0 x1 x2` |
| | `poisson y 0 x1 x2 ; S` |

Estimates a poisson regression. The dependent variable is taken to represent the occurrence of events of some sort, and must take on only non-negative integer values.

If a discrete random variable $Y$ follows the Poisson distribution, then

$$\Pr(Y = y) = \frac{e^{-v} v^y}{y!}$$

for $y = 0, 1, 2, \ldots$. The mean and variance of the distribution are both equal to $v$. In the Poisson regression model, the parameter $v$ is represented as a function of one or more independent variables. The most common version (and the only one supported by gretl) has

$$v = \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots)$$

or in other words the log of $v$ is a linear function of the independent variables.

Optionally, you may add an "offset" variable to the specification. This is a scale variable, the log of which is added to the linear regression function (implicitly, with a coefficient of 1.0). This makes sense if you expect the number of occurrences of the event in question to be proportional, other things equal, to some known factor. For example, the number of traffic accidents might be supposed to be proportional to traffic volume, other things equal, and in that case traffic volume could be specified as an "offset" in a Poisson model of the accident rate. The offset variable must be strictly positive.

By default, standard errors are computed using the negative inverse of the Hessian. If the `--robust` flag is given, then QML or Huber–White standard errors are calculated instead. In this case the estimated covariance matrix is a "sandwich" of the inverse of the estimated Hessian and the outer product of the gradient.

See also negbin.

Menu path: /Model/Nonlinear models/Count data...

**print**

| | |
|---|---|
| Variants: | `print` *varlist* |
| | `print` |
| | `print` *object-name* |
| | `print` *string-literal* |
| Options: | `--byobs` (by observations) |
| | `--no-dates` (use simple observation numbers) |
| Examples: | `print x1 x2 --byobs` |
| | `print my_matrix` |
| | `print "This is a string"` |

If *varlist* is given, prints the values of the specified series, or if no argument is given, prints the values of all series in the current dataset. If the `--byobs` flag is added the data are printed by observation, otherwise they are printed by variable. When printing by observation, the default is to show the date (with time-series data) or the observation marker string (if any) at the start of each line. The `--no-dates` option suppresses the printing of dates or markers; a simple observation number is shown instead.

Besides printing series, you may give the name of a (single) matrix or scalar variable for printing. Or you may give a literal string argument, enclosed in double quotes, to be printed as is. In these case the option flags are not applicable.

Note that you can gain greater control over the printing format (and so, for example, expose a greater number of digits than are shown by default) by using printf.

Menu path: /Data/Display values

**printf**

  Arguments:    *format* , *args*

Prints scalar values, series, matrices, or strings under the control of a format string (providing a subset of the `printf()` statement in the C programming language). Recognized numeric formats are %e, %E, %f, %g, %G and %d, in each case with the various modifiers available in C. Examples: the format `%.10g` prints a value to 10 significant figures; `%12.6f` prints a value to 6 decimal places, with a width of 12 characters. The format `%s` should be used for strings.

The format string itself must be enclosed in double quotes. The values to be printed must follow the format string, separated by commas. These values should take the form of either (a) the names of variables, (b) expressions that are valid for the `genr` command, or (c) the special functions `varname()` or `date()`. The following example prints the values of two variables plus that of a calculated expression:

```
ols 1 0 2 3
scalar b = $coeff[2]
scalar se_b = $stderr[2]
printf "b = %.8g, standard error %.8g, t = %.4f\n",
  b, se_b, b/se_b
```

The next lines illustrate the use of the varname and date functions, which respectively print the name of a variable, given its ID number, and a date string, given a 1-based observation number.

```
printf "The name of variable %d is %s\n", i, varname(i)
printf "The date of observation %d is %s\n", j, date(j)
```

If a matrix argument is given in association with a numeric format, the entire matrix is printed using the specified format for each element. The same applies to series, except that the range of values printed is governed by the current sample setting.

The maximum length of a format string is 127 characters. The escape sequences \n (newline), \t (tab), \v (vertical tab) and \\ (literal backslash) are recognized. To print a literal percent sign, use %%.

As in C, numerical values that form part of the format (width and or precision) may be given directly as numbers, as in `%10.4f`, or they may be given as variables. In the latter case, one puts asterisks into the format string and supplies corresponding arguments in order. For example,

```
scalar width = 12
scalar precision = 6
printf "x = %*.*f\n", width, precision, x
```

**probit**

  Arguments:    *depvar indepvars*
  Options:        `--robust` (robust standard errors)
                `--vcv` (print covariance matrix)
                `--verbose` (print details of iterations)
                `--p-values` (show p-values instead of slopes)

If the dependent variable is a binary variable (all values are 0 or 1) maximum likelihood estimates of the coefficients on *indepvars* are obtained via the Newton–Raphson method.  As the model is nonlinear the slopes depend on the values of the independent variables. By default the slopes with respect to each of the independent variables are calculated (at the means of those variables) and these slopes replace the usual p-values in the regression output. This behavior can be suppressed my giving the `--p-values` option.  The chi-square statistic tests the null hypothesis that all coefficients are zero apart from the constant.

By default, standard errors are computed using the negative inverse of the Hessian.  If the `--robust` flag is given, then QML or Huber–White standard errors are calculated instead. In this case the estimated covariance matrix is a "sandwich" of the inverse of the estimated Hessian and the outer product of the gradient. See chapter 10 of Davidson and MacKinnon for details.

If the dependent variable is not binary but is discrete, then Ordered Probit estimates are obtained. (If the variable selected as dependent is not discrete, an error is flagged.)

Probit for analysis of proportions is not implemented in gretl at this point.

Menu path: /Model/Nonlinear models/Probit


### pvalue

| | |
|---|---|
| Arguments: | *dist* [ *params* ] *xval* |
| Examples: | `pvalue z zscore` |
| | `pvalue t 25 3.0` |
| | `pvalue X 3 5.6` |
| | `pvalue F 4 58 fval` |
| | `pvalue G shape scale x` |
| | `pvalue B bprob 10 6` |
| | `pvalue P lambda x` |
| | `pvalue W shape scale x` |

Computes the area to the right of *xval* in the specified distribution (`z` for Gaussian, `t` for Student's *t*, `X` for chi-square, `F` for *F*, `G` for gamma, `B` for binomial, `P` for Poisson, or `W` for Weibull).

Depending on the distribution, the following information must be given, before the *xval*: for the *t* and chi-square distributions, the degrees of freedom; for *F*, the numerator and denominator degrees of freedom; for gamma, the shape and scale parameters; for the binomial distribution, the "success" probability and the number of trials; for the Poisson distribution, the parameter $\lambda$ (which is both the mean and the variance); and for the Weibull distribution, shape and scale parameters. As shown in the examples above, the numerical parameters may be given in numeric form or as the names of variables.

The parameters for the gamma distribution are sometimes given as mean and variance rather than shape and scale.  The mean is the product of the shape and the scale; the variance is the product of the shape and the square of the scale.  So the scale may be found as the variance divided by the mean, and the shape as the mean divided by the scale.

Menu path: /Tools/P-value finder


### qlrtest

For a model estimated on time-series data via OLS, performs the Quandt likelihood ratio (QLR) test for a structural break at an unknown point in time, with 15 percent trimming at the beginning and end of the sample period.

For each potential break point within the central 70 percent of the observations, a Chow test is performed (see chow). The QLR test statistic is the maximum of the *F* values from these tests. It follows a non-standard distribution, the critical values of which are taken from Stock and Watson (2003). If the QLR statistic exceeds the critical value at the chosen level of significance, one can infer that the parameters of the model are not constant. This statistic can be used to

detect forms of instability other than a single discrete break (such as multiple breaks or a slow drifting of the parameters).

Menu path: Model window, /Tests/QLR test

## qqplot

| Variants: | `qqplot` *y* |
|---|---|
| | `qqplot` *y x* |
| Options: | `--z-scores` (see below) |
| | `--raw` (see below) |

Given just one series argument, displays a plot of the empirical quantiles of the selected series (given by name or ID number) against the quantiles of the normal distribution. The series must include at least 20 valid observations in the current sample range. By default the empirical quantiles are plotted against quantiles of the normal distribution having the same mean and variance as the sample data, but two alternatives are available: if the `--z-scores` option is given the data are standardized, while if the `--raw` option is given the "raw" empirical quantiles are plotted against the quantiles of the standard normal distribution.

Given two series arguments, *y* and *x*, displays a plot of the empirical quantiles of *y* against those of *x*. The data values are not standardized.

Menu path: /Variable/Normal Q-Q plot

Menu path: /View/Graph specified vars/Q-Q plot

## quantreg

| Arguments: | *tau depvar indepvars* |
|---|---|
| Options: | `--robust` (robust standard errors) |
| | `--intervals`[=*level*] (compute confidence intervals) |
| | `--vcv` (print covariance matrix) |
| | `--quiet` (suppress printing of results) |
| Examples: | `quantreg 0.25 y 0 xlist` |
| | `quantreg 0.5 y 0 xlist --intervals` |
| | `quantreg 0.5 y 0 xlist --intervals=.95` |
| | `quantreg tauvec y 0 xlist --robust` |
| | See also `mrw_qr.inp` |

Quantile regression. The first argument, *tau*, is the conditional quantile for which estimates are wanted. It may be given either as a numerical value or as the name of a pre-defined scalar variable; the value must be in the range 0.01 to 0.99. (Alternatively, a vector of values may be given for *tau*; see below for details.) The second and subsequent arguments compose a regression list on the same pattern as ols.

Without the `--intervals` option, standard errors are printed for the quantile estimates. By default, these are computed according to the asymptotic formula given by Koenker and Bassett (1978), but if the `--robust` option is given, standard errors that are robust with respect to heteroskedasticity are calculated using the method of Koenker and Zhao (1994).

When the `--intervals` option is chosen, confidence intervals are given for the parameter estimates instead of standard errors. These intervals are computed using the rank inversion method, and in general they are asymmetrical about the point estimates. The specifics of the calculation are inflected by the `--robust` option: without this, the intervals are computed on the assumption of IID errors (Koenker, 1994); with it, they use the robust estimator developed by Koenker and Machado (1999).

By default, 90 percent confidence intervals are produced. You can change this by appending a confidence level (expressed as a decimal fraction) to the intervals option, as in `--intervals=0.95`.

Vector-valued *tau*: instead of supplying a scalar, you may give the name of a pre-defined matrix. In this case estimates are computed for all the given *tau* values and the results are printed in a special format, showing the sequence of quantile estimates for each regressor in turn.

Menu path: /Model/Robust estimation/Quantile regression

### quit

Exits from the program, giving you the option of saving the output from the session on the way out.

Menu path: /File/Exit

### rename

  Arguments:    *series newname*

Changes the name of *series* (identified by name or ID number) to *newname*. The new name must be of 15 characters maximum, must start with a letter, and must be composed of only letters, digits, and the underscore character.

Menu path: /Variable/Edit attributes

Other access: Main window pop-up menu (single selection)

### reset

  Options:    `--quiet` (don't print the auxiliary regression)

          `--squares-only` (compute the test using only the squares)

          `--cubes-only` (compute the test using only the cubes)

Must follow the estimation of a model via OLS. Carries out Ramsey's RESET test for model specification (non-linearity) by adding the square and/or the cube of the fitted values to the regression and calculating the $F$ statistic for the null hypothesis that the parameters on the added terms are zero.

Both the square and the cube are added, unless one of the options `--squares-only` or `--cubes-only` is given.

Menu path: Model window, /Tests/Ramsey's RESET

### restrict

  Options:    `--quiet` (don't print restricted estimates)

          `--silent` (don't print anything)

          `--wald` (system estimators only – see below)

          `--bootstrap` (bootstrap the test if possible)

          `--full` (OLS and VECMs only, see below)

Imposes a set of (usually linear) restrictions on either (a) the model last estimated or (b) a system of equations previously defined and named. In all cases the set of restrictions should be started with the keyword "restrict" and terminated with "end restrict".

In the single equation case the restrictions are always implicitly to be applied to the last model, and they are evaluated as soon as the `restrict` block is closed.

In the case of a system of equations (defined via the system command), the initial "restrict" may be followed by the name of a previously defined system of equations. If this is omitted and the last model was a system then the restrictions are applied to the last model. By default the restrictions are evaluated when the system is next estimated, using the estimate command. But if the `--wald` option is given the restriction is tested right away, via a Wald chi-square test on the covariance matrix. Note that this option will produce an error if a system has been defined but not yet estimated.

Depending on the context, the restrictions to be tested may be expressed in various ways. The simplest form is as follows: each restriction is given as an equation, with a linear combination of parameters on the left and a scalar value to the right of the equals sign (either a numerical constant or the name of a scalar variable).

In the single-equation case, parameters may be referenced in the form b[$i$], where $i$ represents the position in the list of regressors (starting at 1), or b[*varname*], where *varname* is the name of the regressor in question. In the system case, parameters are referenced using b plus two numbers in square brackets. The leading number represents the position of the equation within the system and the second number indicates position in the list of regressors. For example b[2,1] denotes the first parameter in the second equation, and b[3,2] the second parameter in the third equation. The b terms in the equation representing a restriction may be prefixed with a numeric multiplier, for example 3.5*b[4].

Here is an example of a set of restrictions for a previously estimated model:

```
restrict
 b[1] = 0
 b[2] - b[3] = 0
 b[4] + 2*b[5] = 1
end restrict
```

And here is an example of a set of restrictions to be applied to a named system. (If the name of the system does not contain spaces, the surrounding quotes are not required.)

```
restrict "System 1"
 b[1,1] = 0
 b[1,2] - b[2,2] = 0
 b[3,4] + 2*b[3,5] = 1
end restrict
```

In the single-equation case the restrictions are by default evaluated via a Wald test, using the covariance matrix of the model in question. If the original model was estimated via OLS then the restricted coefficient estimates are printed; to suppress this, append the --quiet option flag to the initial restrict command. As an alternative to the Wald test, for models estimated via OLS or WLS only, you can give the --bootstrap option to perform a bootstrapped test of the restriction.

In the system case, the test statistic depends on the estimator chosen: a Likelihood Ratio test if the system is estimated using a Maximum Likelihood method, or an asymptotic $F$-test otherwise.

There are two alternatives to the method of expressing restrictions discussed above. First, a set of $g$ linear restrictions on a $k$-vector of parameters, $\beta$, may be written compactly as $R\beta - q = 0$, where $R$ is an $g \times k$ matrix and $q$ is a $g$-vector. You can specify a restriction by giving the names of pre-defined, conformable matrices to be used as $R$ and $q$, as in

```
restrict
  R = Rmat
  q = qvec
end restrict
```

Secondly, if you wish to test a nonlinear restriction (this is currently available for single-equation models only) you should give the restriction as the name of a function, preceded by "rfunc = ", as in

```
restrict
  rfunc = myfunction
end restrict
```

The constraint function should take a single const matrix argument; this will be automatically filled out with the parameter vector. And it should return a vector which is zero under the null

hypothesis, non-zero otherwise. The length of the vector is the number of restrictions. This function is used as a "callback" by gretl's numerical Jacobian routine, which calculates a Wald test statistic via the delta method.

Here is a simple example of a function suitable for testing one nonlinear restriction, namely that two pairs of parameter values have a common ratio.

```
function matrix restr (const matrix b)
  matrix v = b[1]/b[2] - b[4]/b[5]
  return v
end function
```

On successful completion of the `restrict` command the accessors `$test` and `$pvalue` give the test statistic and its p-value.

When testing restrictions on a single-equation model estimated via OLS, or on a VECM, the `--full` option can be used to set the restricted estimates as the "last model" for the purposes of further testing or the use of accessors such as `$coeff` and `$vcv`. Note that some special considerations apply in the case of testing restrictions on Vector Error Correction Models. Please see the *Gretl User's Guide* for details.

Menu path: Model window, /Tests/Linear restrictions

### rmplot

Argument:   *series*
Options:    `--trim` (see below)
            `--quiet` (suppress printed output)

Range–mean plot: this command creates a simple graph to help in deciding whether a time series, $y(t)$, has constant variance or not. We take the full sample t=1,...,T and divide it into small subsamples of arbitrary size $k$. The first subsample is formed by $y(1),...,y(k)$, the second is $y(k+1), ..., y(2k)$, and so on. For each subsample we calculate the sample mean and range (= maximum minus minimum), and we construct a graph with the means on the horizontal axis and the ranges on the vertical. So each subsample is represented by a point in this plane. If the variance of the series is constant we would expect the subsample range to be independent of the subsample mean; if we see the points approximate an upward-sloping line this suggests the variance of the series is increasing in its mean; and if the points approximate a downward sloping line this suggests the variance is decreasing in the mean.

Besides the graph, gretl displays the means and ranges for each subsample, along with the slope coefficient for an OLS regression of the range on the mean and the p-value for the null hypothesis that this slope is zero. If the slope coefficient is significant at the 10 percent significance level then the fitted line from the regression of range on mean is shown on the graph. The $t$-statistic for the null, and the corresponding p-value, are recorded and may be retrieved using the accessors `$test` and `$pvalue` respectively.

If the `--trim` option is given, the minimum and maximum values in each sub-sample are discarded before calculating the mean and range. This makes it less likely that outliers will distort the analysis.

If the `--quiet` option is given, no graph is shown and no output is printed; only the $t$-statistic and p-value are recorded.

Menu path: /Variable/Range-mean graph

### run

Argument:   *filename*

Execute the commands in *filename* then return control to the interactive prompt. This command is intended for use with the command-line program gretlcli, or at the "gretl console" in the GUI program.

See also include.

Menu path: Run icon in script window

### runs

Argument: *series*

Options: `--difference` (use first difference of variable)

     `--equal` (positive and negative values are equiprobable)

Carries out the nonparametric "runs" test for randomness of the specified *series*, where runs are defined as sequences of consecutive positive or negative values. If you want to test for randomness of deviations from the median, for a variable named `x1` with a non-zero median, you can do the following:

```
series signx1 = x1 - median(x1)
runs signx1
```

If the `--difference` option is given, the variable is differenced prior to the analysis, hence the runs are interpreted as sequences of consecutive increases or decreases in the value of the variable.

If the `--equal` option is given, the null hypothesis incorporates the assumption that positive and negative values are equiprobable, otherwise the test statistic is invariant with respect to the "fairness" of the process generating the sequence, and the test focuses on independence alone.

Menu path: /Tools/Nonparametric tests

### scatters

Arguments: *yvar ; xvars* or *yvars ; xvar*

Options: `--with-lines` (create line graphs)

     `--matrix=`*name* (plot columns of named matrix)

     `--output=`*filename* (send output to specified file)

Examples: `scatters 1 ; 2 3 4 5`

     `scatters 1 2 3 4 5 6 ; 7`

     `scatters y1 y2 y3 ; x --with-lines`

Generates pairwise graphs of *yvar* against all the variables in *xvars*, or of all the variables in *yvars* against *xvar*. The first example above puts variable 1 on the $y$-axis and draws four graphs, the first having variable 2 on the $x$-axis, the second variable 3 on the $x$-axis, and so on. The second example plots each of variables 1 through 6 against variable 7 on the $x$-axis. Scanning a set of such plots can be a useful step in exploratory data analysis. The maximum number of plots is 16; any extra variable in the list will be ignored.

By default the graphs are scatterplots, but if you give the `--with-lines` flag they will be line graphs.

For details on usage of the `--output` option, please see the gnuplot command.

If a named matrix is specified as the data source the $x$ and $y$ lists should be given as 1-based column numbers; or alternatively, if no such numbers are given, all the columns are plotted against time or an index variable.

Menu path: /View/Multiple graphs

### sdiff

Argument: *varlist*

The seasonal difference of each variable in *varlist* is obtained and the result stored in a new variable with the prefix `sd_`. This command is available only for seasonal time series.

Menu path: /Add/Seasonal differences of selected variables

**set**

| Variants: | `set` *variable value* |
|---|---|
| | `set --to-file=`*filename* |
| | `set --from-file=`*filename* |
| | `set stopwatch` |
| | `set` |
| Examples: | `set svd on` |
| | `set csv_delim tab` |
| | `set horizon 10` |
| | `set --to-file=mysettings.inp` |

The most common use of this command is the first variant shown above, where it is used to set the value of a selected program parameter. This is discussed in detail below. The other uses are: with `--to-file`, to write a script file containing all the current parameter settings; with `--from-file` to read a script file containing parameter settings and apply them to the current session; with `stopwatch` to zero the gretl "stopwatch" which can be used to measure CPU time (see the entry for the `$stopwatch` accessor in the gretl function reference); or, if the word `set` is given alone, to print the current settings.

Values set via this comand remain in force for the duration of the gretl session unless they are changed by a further call to `set`. The parameters that can be set in this way are enumerated below. Note that the settings of `hc_version`, `hac_lag` and `hac_kernel` are used when the `--robust` option is given to an estimation command.

The available settings are grouped under the following categories: program interaction and behavior, numerical methods, random number generation, robust estimation, filtering, time series estimation, and interaction with GNU R.

*Program interaction and behavior*

These settings are used for controlling various aspects of the way gretl interacts with the user.

- `csv_delim`: either `comma` (the default), `space`, `tab` or `semicolon`. Sets the column delimiter used when saving data to file in CSV format.

- `csv_na`: string to represent missing values when writing data to file in CSV format. Maximum 7 characters; the default is `NA`.

- `echo`: `off` or `on` (the default). Suppress or resume the echoing of commands in gretl's output.

- `force_decpoint`: `on` or `off` (the default). Force gretl to use the decimal point character, in a locale where another character (most likely the comma) is the standard decimal separator.

- `halt_on_error`: `off` or `on` (the default). By default, when an error is encountered in the course of executing a script, execution is halted (and if the command-line program is operating in batch mode, it exits with a non-zero return status). You can force gretl to continue on error by setting `halt_on_error` to `off` (or by setting the environment variable `GRETL_KEEP_GOING` to 1). If an error occurs while "compiling" a loop or user-defined function, however, execution is halted regardless.

- `loop_maxiter`: one positive integer value. Sets the maximum number of iterations that a `while` loop is allowed before halting (see loop). Note that this setting only affects the `while` variant, as its purpose is to forcibly break potentially infinite loops.

- `max_verbose`: `on` or `off` (the default). Toggles verbose output for the `BFGSmax` function (see the User's Guide for details).

- `messages`: `off` or `on` (the default). Suppress or resume the printing of non-error messages associated with various commands, for example when a new variable is generated or when the sample range is changed.

- **warnings:** `off` or `on` (the default). Suppress or resume the printing of warning messages issued when arithmetical operations produce non-finite values.

- **debug:** 1, 2 or 0 (the default). This is for use with user-defined functions. Setting `debug` to 1 is equivalent to turning `messages` on within all such functions; setting this variable to 2 has the additional effect of turning on `max_verbose` within all functions.

- **shell_ok:** `on` or `off` (the default). Enable launching external programs from gretl via the system shell. This is disabled by default for security reasons, and can only be enabled via the graphical user interface (Tools/Preferences/General). However, once set to on, this setting will remain active for future sessions until explicitly disabled.

- **shelldir:** *path*. Sets the current working directory for shell commands.

- **use_cwd:** `on` or `off` (the default). This setting affects the behavior of the outfile and store commands, which write external files. Normally, the file will be written in the user's default data directory; if `use_cwd` is on, on the contrary, the file will be created in the working directory when gretl was started.

- **bfgs_verbskip:** one integer. This setting affects the behavior of the `--verbose` option to those commands that use BFGS as an optimization algorithm and is used to compact output. if `bfgs_verbskip` is set to, say, 3, then the `--verbose` switch will only print iterations 3, 6, 9 and so on.

- **skip_missing:** `on` (the default) or `off`. Controls gretl's behavior when contructing a matrix from data series: the default is to skip data rows that contain one or more missing values but if `skip_missing` is set `off` missing values are converted to NaNs.

- **matrix_mask:** the name of a series, or the keyword `null`. Offers greater control than `skip_missing` when constructing matrices from series: the data rows selected for matrices are those with non-zero (and non-missing) values in the specified series. The selected mask remains in force until it is replaced, or removed via the `null` keyword.

*Numerical methods*

These settings are used for controlling the numerical algorithms that gretl uses for estimation.

- **optimizer:** either `auto` (the default), BFGS or `newton`. Sets the optimization algorithm used for various ML estimators, in cases where both BFGS and Newton–Raphson are applicable. The default is to use Newton–Raphson where an analytical Hessian is available, otherwise BFGS.

- **bhhh_maxiter:** one integer, the maximum number of iterations for gretl's internal BHHH routine, which is used in the `arma` command for conditional ML estimation. If convergence is not achieved after `bhhh_maxiter`, the program returns an error. The default is set at 500.

- **bhhh_toler:** one floating point value, or the string `default`. This is used in gretl's internal BHHH routine to check if convergence has occurred. The algorithm stops iterating as soon as the increment in the log-likelihood between iterations is smaller than `bhhh_toler`. The default value is $1.0E-06$; this value may be re-established by typing `default` in place of a numeric value.

- **bfgs_maxiter:** one integer, the maximum number of iterations for gretl's BFGS routine, which is used for `mle`, `gmm` and several specific estimators. If convergence is not achieved in the specified number of iterations, the program returns an error. The default value depends on the context, but is typically of the order of 500.

- **bfgs_toler:** one floating point value, or the string `default`. This is used in gretl's BFGS routine to check if convergence has occurred. The algorithm stops as soon as the relative improvement in the objective function between iterations is smaller than `bfgs_toler`. The default value is the machine precision to the power 3/4; this value may be re-established by typing `default` in place of a numeric value.

- **bfgs_maxgrad**: one floating point value. This is used in gretl's BFGS routine to check if the norm of the gradient is reasonably close to zero when the **bfgs_toler** criterion is met. A warning is printed if the norm of the gradient exceeds 1; an error is flagged if the norm exceeds **bfgs_maxgrad**. At present the default is the permissive value of 5.0.

- **bfgs_richardson**: **on** or **off** (the default). Use Richardson extrapolation when computing numerical derivatives in the context of BFGS maximization.

- **initvals**: a pre-specified matrix. Allows manual setting of the initial parameter estimates for numerical optimization problems (such as ARMA estimation). For details see the *Gretl User's Guide.*

- **lbfgs**: **on** or **off** (the default). Use the limited-memory version of BFGS (L-BFGS-B) instead of the ordinary algorithm. This may be advantageous when the function to be maximized is not globally concave.

- **lbfgs_mem**: an integer value in the range 3 to 20 (with a default value of 8). This determines the number of corrections used in the limited memory matrix when L-BFGS-B is employed.

- **nls_toler**: a floating-point value (the default is the machine precision to the power 3/4). Sets the tolerance used in judging whether or not convergence has occurred in nonlinear least squares estimation using the nls command.

- **svd**: **on** or **off** (the default). Use SVD rather than Cholesky or QR decomposition in least squares calculations. This option applies to the **mols** function as well as various internal calculations, but not to the regular ols command.

- **fcp**: **on** or **off** (the default). Use the algorithm of Fiorentini, Calzolari and Panattoni rather than native gretl code when computing GARCH estimates.

- **gmm_maxiter**: one integer, the maximum number of iterations for gretl's **gmm** command when in iterated mode (as opposed to one- or two-step). The default value is 250.

- **nadarwat_trim**: one integer, the trim parameter used in the **nadarwat** function.

*Random number generation*

- **seed**: an unsigned integer. Sets the seed for the pseudo-random number generator. By default this is set from the system time; if you want to generate repeatable sequences of random numbers you must set the seed manually.

- **RNG**: either **MT** or **SFMT** (the default). Switches between the default random number generator, namely the SIMD-oriented Fast Mersenne Twister (SFMT), and the Mersenne Twister of 2002 as implemented in GLib (MT). SFMT is faster and has better distributional properties but MT was gretl's RNG up to version 1.9.3.

- **normal_rand**: **ziggurat** (the default) or **box-muller**. Sets the method for generating random normal samples based on uniform input.

*Robust estimation*

- **bootrep**: an integer. Sets the number of replications for the restrict command with the **--bootstrap** option.

- **garch_vcv**: **unset**, **hessian**, **im** (information matrix) , **op** (outer product matrix), **qml** (QML estimator), **bw** (Bollerslev–Wooldridge). Specifies the variant that will be used for estimating the coefficient covariance matrix, for GARCH models. If **unset** is given (the default) then the Hessian is used unless the "robust" option is given for the garch command, in which case QML is used.

- **arma_vcv**: **hessian** (the default) or **op** (outer product matrix). Specifies the variant to be used when computing the covariance matrix for ARIMA models.

- `force_hc`: `off` (the default) or `on`. By default, with time-series data and when the `--robust` option is given with `ols`, the HAC estimator is used. If you set `force_hc` to "on", this forces calculation of the regular Heteroskedasticity Consistent Covariance Matrix (HCCM), which does not take autocorrelation into account. Note that when a VAR is estimated with the `--robust` option, the default is regular HCCM, but the `--robust-hac` option is available to produce a HAC estimator.

- `hac_lag`: `nw1` (the default), `nw2`, `nw3` or an integer. Sets the maximum lag value or bandwidth, $p$, used when calculating HAC (Heteroskedasticity and Autocorrelation Consistent) standard errors using the Newey-West approach, for time series data. `nw1` and `nw2` represent two variant automatic calculations based on the sample size, $T$: for nw1, $p = 0.75 \times T^{1/3}$, and for nw2, $p = 4 \times (T/100)^{2/9}$. `nw3` calls for data-based bandwidth selection. See also `qs_bandwidth` and `hac_prewhiten` below.

- `hac_kernel`: `bartlett` (the default), `parzen`, or `qs` (Quadratic Spectral). Sets the kernel, or pattern of weights, used when calculating HAC standard errors.

- `hac_prewhiten`: `on` or `off` (the default). Use Andrews-Monahan prewhitening and recoloring when computing HAC standard errors. This also implies use of data-based bandwidth selection.

- `hc_version`: 0 (the default), 1, 2, 3 or 3a. Sets the variant used when calculating Heteroskedasticity Consistent standard errors with cross-sectional data. The first four options correspond to the HC0, HC1, HC2 and HC3 discussed by Davidson and MacKinnon in *Econometric Theory and Methods*, chapter 5. HC0 produces what are usually called "White's standard errors". Variant 3a is the MacKinnon–White "jackknife" procedure.

- `pcse`: `off` (the default) or `on`. By default, when estimating a model using pooled OLS on panel data with the `--robust` option, the Arellano estimator is used for the covariance matrix. If you set `pcse` to "on", this forces use of the Beck and Katz Panel Corrected Standard Errors (which do not take autocorrelation into account).

- `qs_bandwidth`: Bandwidth for HAC estimation in the case where the Quadratic Spectral kernel is selected. (Unlike the Bartlett and Parzen kernels, the QS bandwidth need not be an integer.)

*Time series*

- `horizon`: one integer (the default is based on the frequency of the data). Sets the horizon for impulse responses and forecast variance decompositions in the context of vector autoregressions.

- `vecm_norm`: `phillips` (the default), `diag`, `first` or `none`. Used in the context of VECM estimation via the vecm command for identifying the cointegration vectors. See the the *Gretl User's Guide* for details.

*Interaction with R*

- `R_lib`: `on` (the default) or `off`. When sending instructions to be executed by R, use the R shared library by preference to the R executable, if the library is available.

- `R_functions`: `off` (the default) or `on`. Recognize functions defined in R as if they were native functions (the namespace prefix "`R.`" is required). See the *Gretl User's Guide* for details on this and the previous item.

### setinfo

| | |
|---|---|
| Argument: | *series* |
| Options: | `--description=`*string* (set description) |
| | `--graph-name=`*string* (set graph name) |
| | `--discrete` (mark series as discrete) |
| | `--continuous` (mark series as continuous) |
| Examples: | `setinfo x1 --description="Description of x1"` |
| | `setinfo y --graph-name="Some string"` |
| | `setinfo z --discrete` |

Sets up to three attributes of *series*, given by name or ID number, as follows.

If the `--description` flag is given followed by a string in double quotes, that string is used to set the variable's descriptive label. This label is shown in response to the labels command, and is also shown in the main window of the GUI program.

If the `--graph-name` flag is given followed by a quoted string, that string will be used in place of the variable's name in graphs.

If one or other of the `--discrete` or `--continuous` option flags is given, the variable's numerical character is set accordingly. The default is to treat all series as continuous; setting a series as discrete affects the way the variable is handled in frequency plots.

Menu path: /Variable/Edit attributes

Other access: Main window pop-up menu

### setobs

| | |
|---|---|
| Variants: | `setobs` *periodicity startobs* |
| | `setobs` *unitvar timevar* `--panel-vars` |
| | `setobs --labels=`*filename* |
| Options: | `--cross-section` (interpret as cross section) |
| | `--time-series` (interpret as time series) |
| | `--stacked-cross-section` (interpret as panel data) |
| | `--stacked-time-series` (interpret as panel data) |
| | `--panel-vars` (use index variables, see below) |
| Examples: | `setobs 4 1990:1 --time-series` |
| | `setobs 12 1978:03` |
| | `setobs 1 1 --cross-section` |
| | `setobs 20 1:1 --stacked-time-series` |
| | `setobs unit year --panel-vars` |

In its primary use, this command forces the program to interpret the current data set as having a specified structure. A secondary use, with the `--labels` flag, is to add observation labels to a dataset.

In the first form of the command the *periodicity*, which must be an integer, represents frequency in the case of time-series data (1 = annual; 4 = quarterly; 12 = monthly; 52 = weekly; 5, 6, or 7 = daily; 24 = hourly). In the case of panel data the periodicity means the number of lines per data block: this corresponds to the number of cross-sectional units in the case of stacked cross-sections, or the number of time periods in the case of stacked time series. In the case of simple cross-sectional data the periodicity should be set to 1.

The starting observation represents the starting date in the case of time series data. Years may be given with two or four digits; subperiods (for example, quarters or months) should be separated from the year with a colon. In the case of panel data the starting observation should be given as 1:1; and in the case of cross-sectional data, as 1. Starting observations for daily or weekly data should be given in the form YY/MM/DD or YYYY/MM/DD (or simply as 1 for

undated data).

The second form of the command (which requires the `--panel-vars` flag) may be used to impose a panel interpretation when the data set contains variables that uniquely identify the cross-sectional units and the time periods. The data set will be sorted as stacked time series, by ascending values of the units variable, *unitvar*.

If no explicit option flag is given to indicate the structure of the data the program will attempt to guess the structure from the information given.

In the third form of the comman, with the `--labels` flag, the effect is not to set the interpretation of the data structure, but rather to add labels for the observations. The required *filename* parameter must give the name of a plain text file that contains one label per line, with enough lines to label all the observations.

Menu path: /Data/Dataset structure

Menu path: /Data/Observation markers

### setmiss

| | |
|---|---|
| Arguments: | *value* [ *varlist* ] |
| Examples: | `setmiss -1` |
| | `setmiss 100 x2` |

Get the program to interpret some specific numerical data value (the first parameter to the command) as a code for "missing", in the case of imported data. If this value is the only parameter, as in the first example above, the interpretation will be applied to all series in the data set. If *value* is followed by a list of variables, by name or number, the interpretation is confined to the specified variable(s). Thus in the second example the data value 100 is interpreted as a code for "missing", but only for the variable `x2`.

Menu path: /Data/Set missing value code

### shell

| | |
|---|---|
| Argument: | *shellcommand* |
| Examples: | `!  ls -al` |
| | `!  notepad` |
| | `launch notepad` |

A `!`, or the keyword `launch`, at the beginning of a command line is interpreted as an escape to the user's shell. Thus arbitrary shell commands can be executed from within gretl. When `!` is used, the external command is executed synchronously. That is, gretl waits for it to complete before proceeding. If you want to start another program from within gretl and not wait for its completion (asynchronous operation), use `launch` instead.

For reasons of security this facility is not enabled by default. To activate it, check the box titled "Allow shell commands" under the File, Preferences menu in the GUI program. This also makes shell commands available in the command-line program (and is the only way to do so).

**smpl**

| | |
|---|---|
| Variants: | smpl *startobs endobs* |
| | smpl *+i -j* |
| | smpl *dumvar* --dummy |
| | smpl *condition* --restrict |
| | smpl --no-missing [ *varlist* ] |
| | smpl --contiguous [ *varlist* ] |
| | smpl *n* --random |
| | smpl full |
| Options: | --dummy (argument is a dummy variable) |
| | --restrict (apply boolean restriction) |
| | --replace (replace any existing boolean restriction) |
| | --no-missing (restrict to valid observations) |
| | --contiguous (see below) |
| | --random (form random sub-sample) |
| | --balanced (panel data: try to retain balanced panel) |
| Examples: | smpl 3 10 |
| | smpl 1960:2 1982:4 |
| | smpl +1 -1 |
| | smpl x > 3000 --restrict |
| | smpl y > 3000 --restrict --replace |
| | smpl 100 --random |

Resets the sample range. The new range can be defined in several ways. In the first alternate form (and the first two examples) above, *startobs* and *endobs* must be consistent with the periodicity of the data. Either one may be replaced by a semicolon to leave the value unchanged. In the second form, the integers *i* and *j* (which may be positive or negative, and should be signed) are taken as offsets relative to the existing sample range. In the third form *dummyvar* must be an indicator variable with values 0 or 1 at each observation; the sample will be restricted to observations where the value is 1. The fourth form, using --restrict, restricts the sample to observations that satisfy the given Boolean condition (which is specified according to the syntax of the genr command).

With the --no-missing form, if *varlist* is specified observations are selected on condition that all variables in *varlist* have valid values at that observation; otherwise, if no *varlist* is given, observations are selected on condition that *all* variables have valid (non-missing) values.

The --contiguous form of smpl is intended for use with time series data. The effect is to trim any observations at the start and end of the current sample range that contain missing values (either for the variables in *varlist*, or for all data series if no *varlist* is given). Then a check is performed to see if there are any missing values in the remaining range; if so, an error is flagged.

With the --random flag, the specified number of cases are selected from the current dataset at random (without replacement). If you wish to be able to replicate this selection you should set the seed for the random number generator first (see the set command).

The final form, smpl full, restores the full data range.

Note that sample restrictions are, by default, cumulative: the baseline for any smpl command is the current sample. If you wish the command to act so as to replace any existing restriction you can add the option flag --replace to the end of the command. (But this option is not compatible with the --contiguous option.)

The internal variable obs may be used with the --restrict form of smpl to exclude particular observations from the sample. For example

```
smpl obs!=4 --restrict
```

will drop just the fourth observation. If the data points are identified by labels,

```
smpl obs!="USA" --restrict
```

will drop the observation with label "USA".

One point should be noted about the `--dummy`, `--restrict` and `--no-missing` forms of `smpl`: "structural" information in the data file (regarding the time series or panel nature of the data) is likely to be lost when this command is issued. You may reimpose structure with the setobs command. A related option, for use with panel data, is the `--balanced` flag: this requests that a balanced panel is reconstituted after sub-sampling, via the insertion of "missing rows" if need be. But note that it is not always possible to comply with this request.

Please see the *Gretl User's Guide* for further details.

Menu path: /Sample

### spearman

Arguments: *var1 var2*
Option: `--verbose` (print ranked data)

Prints Spearman's rank correlation coefficient for the series *var1* and *var2*. The variables do not have to be ranked manually in advance; the function takes care of this.

The automatic ranking is from largest to smallest (i.e. the largest data value gets rank 1). If you need to invert this ranking, create a new variable which is the negative of the original. For example:

```
series altx = -x
spearman altx y
```

Menu path: /Model/Robust estimation/Rank correlation

### sprintf

Arguments: *stringvar format* , *args*

This command works exactly like the printf command, printing the given arguments under the control of the format string, except that the result is written into the named string, *stringvar*.

To print the string variable, give its name preceded by @, as in

```
sprintf myvar "%g", x
print "myvar has value @myvar"
```

### square

Argument: *varlist*
Option: `--cross` (generate cross-products as well as squares)

Generates new series which are squares of the series in *varlist* (plus cross-products if the `--cross` option is given). For example, `square x y` will generate `sq_x = x` squared, `sq_y = y` squared and (optionally) `x_y = x` times y. If a particular variable is a dummy variable it is not squared because we will get the same variable.

Menu path: /Add/Squares of selected variables

### sscanf

Arguments: *source* , *format* , *args*

Deprecated command-style version of the sscanf function.

**store**

| | |
|---|---|
| Arguments: | *filename* [ *varlist* ] |
| Options: | `--csv` (use CSV format) |
| | `--omit-obs` (see below, on CSV format) |
| | `--no-header` (see below, on CSV format) |
| | `--gnu-octave` (use GNU Octave format) |
| | `--gnu-R` (use GNU R format) |
| | `--traditional` (use traditional ESL format) |
| | `--gzipped` (apply gzip compression) |
| | `--jmulti` (use JMulti ASCII format) |
| | `--dat` (use PcGive ASCII format) |
| | `--decimal-comma` (use comma as decimal character) |
| | `--database` (use gretl database format) |
| | `--overwrite` (see below, on database format) |
| | `--comment=`*string* (see below) |

Saves either the entire dataset or, if a *varlist* is supplied, a specified subset of the series in the current dataset, to the file given by *filename*.

By default the data are saved in "native" gretl format, but the option flags permit saving in several alternative formats. CSV (Comma-Separated Values) data may be read into spreadsheet programs, and can also be manipulated using a text editor. The formats of Octave, R and PcGive are designed for use with the respective programs. Gzip compression may be useful for large datasets. See the *Gretl User's Guide* for details on the various formats.

The option flags `--omit-obs` and `--no-header` are applicable only when saving data in CSV format. By default, if the data are time series or panel, or if the dataset includes specific observation markers, the CSV file includes a first column identifying the observations (e.g. by date). If the `--omit-obs` flag is given this column is omitted. The `--no-header` flag suppresses the usual printing of the names of the variables at the top of the columns.

The option flag `--decimal-comma` is also confined to the case of saving data in CSV format. The effect of this option is to replace the decimal point with the decimal comma; in addition the column separator is forced to be a semicolon.

The option of saving in gretl database format is intended to help with the construction of large sets of series, possibly having mixed frequencies and ranges of observations. At present this option is available only for annual, quarterly or monthly time-series data. If you save to a file that already exists, the default action is to append the newly saved series to the existing content of the database. In this context it is an error if one or more of the variables to be saved has the same name as a variable that is already present in the database. The `--overwrite` flag has the effect that, if there are variable names in common, the newly saved variable replaces the variable of the same name in the original dataset.

The `--comment` option is available when saving data as a database or in CSV format. The required parameter is a double-quoted one-line string, attached to the option flag with an equals sign. The string is inserted as a comment into the database index file or at the top of the CSV output.

The `store` command behaves in a special manner in the context of a "progressive loop". See the *Gretl User's Guide* for details.

Menu path: /File/Save data; /File/Export data

**summary**

Variants:   `summary` [ *varlist* ]
            `summary --matrix=`*matname*

Options:    `--simple` (basic statistics only)
            `--by=`*byvar* (see below)

In its first form, this command prints summary statistics for the variables in *varlist*, or for all the variables in the data set if *varlist* is omitted. By default, output consists of the mean, standard deviation (sd), coefficient of variation (= sd/mean), median, minimum, maximum, skewness coefficient, and excess kurtosis. If the `--simple` option is given, output is restricted to the mean, minimum, maximum and standard deviation.

If the `--by` option is given (in which case the parameter *byvar* should be the name of a discrete variable), then statistics are printed for sub-samples corresponding to the distinct values taken on by *byvar*. For example, if *byvar* is a (binary) dummy variable, statistics are given for the cases `byvar = 0` and `byvar = 1`.

If the alternative form is given, using a named matrix, then summary statistics are printed for each column of the matrix. The `--by` option is not available in this case.

Menu path: /View/Summary statistics

Other access: Main window pop-up menu

**system**

Variants:   `system method=`*estimator*
            *sysname* `<- system`

Examples:   `"Klein Model 1" <- system`
            `system method=sur`
            `system method=3sls`
            See also `klein.inp`, `kmenta.inp`, `greene14_2.inp`

Starts a system of equations. Either of two forms of the command may be given, depending on whether you wish to save the system for estimation in more than one way or just estimate the system once.

To save the system you should assign it a name, as in the first example (if the name contains spaces it must be surrounded by double quotes). In this case you estimate the system using the estimate command. With a saved system of equations, you are able to impose restrictions (including cross-equation restrictions) using the restrict command.

Alternatively you can specify an estimator for the system using `method=` followed by a string identifying one of the supported estimators: `ols` (Ordinary Least Squares), `tsls` (Two-Stage Least Squares) `sur` (Seemingly Unrelated Regressions), `3sls` (Three-Stage Least Squares), `fiml` (Full Information Maximum Likelihood) or `liml` (Limited Information Maximum Likelihood). In this case the system is estimated once its definition is complete.

An equation system is terminated by the line `end system`. Within the system four sorts of statement may be given, as follows.

- equation: specify an equation within the system. At least two such statements must be provided.

- `instr`: for a system to be estimated via Three-Stage Least Squares, a list of instruments (by variable name or number). Alternatively, you can put this information into the `equation` line using the same syntax as in the tsls command.

- `endog`: for a system of simultaneous equations, a list of endogenous variables. This is primarily intended for use with FIML estimation, but with Three-Stage Least Squares this approach may be used instead of giving an `instr` list; then all the variables not identified as endogenous will be used as instruments.

- `identity`: for use with FIML, an identity linking two or more of the variables in the system. This sort of statement is ignored when an estimator other than FIML is used.

After estimation using the `system` or `estimate` commands the following accessors can be used to retrieve additional information:

- `$uhat`: the matrix of residuals, one column per equation.

- `$yhat`: matrix of fitted values, one column per equation.

- `$coeff`: column vector of coefficients (all the coefficients from the first equation, followed by those from the second equation, and so on).

- `$vcv`: covariance matrix of the coefficients. If there are $k$ elements in the `$coeff` vector, this matrix is $k$ by $k$.

- `$sigma`: cross-equation residual covariance matrix.

- `$sysGamma`, `$sysA` and `$sysB`: structural-form coefficient matrices (see below).

If you want to retrieve the residuals or fitted values for a specific equation as a data series, select a column from the `$uhat` or `$yhat` matrix and assign it to a series, as in

```
series uh1 = $uhat[,1]
```

The structural-form matrices correspond to the following representation of a simultaneous equations model:

$$\Gamma y_t = A y_{t-1} + B x_t + \epsilon_t$$

If there are $n$ endogenous variables and $k$ exogenous variables, $\Gamma$ is an $n \times n$ matrix and $B$ is $n \times k$. If the system contains no lags of the endogenous variables then the $A$ matrix is not present. If the maximum lag of an endogenous regressor is $p$, the $A$ matrix is $n \times np$.

Menu path: /Model/Simultaneous equations

**tabprint**

| | |
|---|---|
| Argument: | [ *-f filename* ] |
| Options: | `--rtf` (Produce RTF instead of LaTeX) |
| | `--complete` (Create a complete document) |
| | `--format="f1\|f2\|f3\|f4"` (Specify a custom format) |

Must follow the estimation of a model. Prints the estimated model in tabular form — by default as LaTeX, but as RTF if the `--rtf` flag is given. If a filename is specified using the `-f` flag output goes to that file, otherwise it goes to a file with a name of the form `model_N.tex` (or `model_N.rtf`), where N is the number of models estimated to date in the current session.

The further options discussed below are available only when printing the model as LaTeX.

If the `--complete` flag is given the LaTeX file is a complete document, ready for processing; otherwise it must be included in a document.

If you wish alter the appearance of the tabular output, you can specify a custom row format using the `--format` flag. The format string must be enclosed in double quotes and must be tied to the flag with an equals sign. The pattern for the format string is as follows. There are four fields, representing the coefficient, standard error, $t$-ratio and p-value respectively. These fields should be separated by vertical bars; they may contain a `printf`-type specification for the formatting of the numeric value in question, or may be left blank to suppress the printing of that column (subject to the constraint that you can't leave all the columns blank). Here are a few examples:

```
--format="%.4f|%.4f|%.4f|%.4f"
--format="%.4f|%.4f|%.3f|"
--format="%.5f|%.4f||%.4f"
--format="%.8g|%.8g||%.4f"
```

The first of these specifications prints the values in all columns using 4 decimal places. The second suppresses the p-value and prints the $t$-ratio to 3 places. The third omits the $t$-ratio. The last one again omits the $t$, and prints both coefficient and standard error to 8 significant figures.

Once you set a custom format in this way, it is remembered and used for the duration of the gretl session. To revert to the default format you can use the special variant `--format=default`.

Menu path: Model window, /LaTeX

### textplot

| Argument: | *varlist* |
|---|---|
| Options: | `--time-series` (plot by observation) |
| | `--one-scale` (force a single scale) |
| | `--tall` (use 40 rows) |

Quick and simple ASCII graphics. Without the `--time-series` flag, *varlist* must contain at least two series, the last of which is taken as the variable for the $x$ axis, and a scatter plot is produced. In this case the `--tall` option may be used to produce a graph in which the $y$ axis is represented by 40 rows of characters (the default is 20 rows).

With the `--time-series`, a plot by observation is produced. In this case the option `--one-scale` may be used to force the use of a single scale; otherwise if *varlist* contains more than one series the data may be scaled. Each line represents an observation, with the data values plotted horizontally.

See also gnuplot.

### tobit

| Arguments: | *depvar indepvars* |
|---|---|
| Options: | `--llimit=`*lval* (specify left bound) |
| | `--rlimit=`*rval* (specify right bound) |
| | `--vcv` (print covariance matrix) |
| | `--robust` (robust standard errors) |
| | `--verbose` (print details of iterations) |

Estimates a Tobit model, which may be appropriate when the dependent variable is "censored". For example, positive and zero values of purchases of durable goods on the part of individual households are observed, and no negative values, yet decisions on such purchases may be thought of as outcomes of an underlying, unobserved disposition to purchase that may be negative in some cases.

By default it is assumed that the dependent variable is censored at zero on the left and is uncensored on the right. However you can use the options `--llimit` and `--rlimit` to specify a different pattern of censoring. Note that if you specify a right bound only, the assumption is then that the dependent variable is uncensored on the left.

The Tobit model is a special case of interval regression, which is supported via the intreg command.

Menu path: /Model/Nonlinear models/Tobit

**tsls**

| | |
|---|---|
| Arguments: | *depvar indepvars* ; *instruments* |
| Options: | `--no-tests` (don't do diagnostic tests) |
| | `--vcv` (print covariance matrix) |
| | `--robust` (robust standard errors) |
| | `--liml` (use Limited Information Maximum Likelihood) |
| | `--gmm` (use the Generalized Method of Moments) |
| Example: | `tsls y1 0 y2 y3 x1 x2 ; 0 x1 x2 x3 x4 x5 x6` |

Computes Instrumental Variables (IV) estimates, by default using two-stage least squares (TSLS) but see below for further options. The dependent variable is *depvar*, *indepvars* is the list of regressors (which is presumed to include at least one endogenous variable); and *instruments* is the list of instruments (exogenous and/or predetermined variables). If the *instruments* list is not at least as long as *indepvars*, the model is not identified.

In the above example, the ys are endogenous and the xs are the exogenous variables. Note that exogenous regressors should appear in both lists.

Output for two-stage least squares estimates includes the Hausman test and, if the model is over-identified, the Sargan over-identification test. In the Hausman test, the null hypothesis is that OLS estimates are consistent, or in other words estimation by means of instrumental variables is not really required. A model of this sort is over-identified if there are more instruments than are strictly required. The Sargan test is based on an auxiliary regression of the residuals from the two-stage least squares model on the full list of instruments. The null hypothesis is that all the instruments are valid, and suspicion is thrown on this hypothesis if the auxiliary regression has a significant degree of explanatory power. For a good explanation of both tests see chapter 8 of Davidson and MacKinnon (2004).

For both TSLS and LIML estimation, an additional test result is shown provided that the model is estimated under the assumption of i.i.d. errors (that is, the `--robust` option is not selected). This is a test for weakness of the instruments. Weak instruments can lead to serious problems in IV regression: biased estimates and/or incorrect size of hypothesis tests based on the covariance matrix, with rejection rates well in excess of the nominal significance level (Stock *et al.*, 2002). The test statistic is the first-stage *F*-test if the model contains just one endogenous regressor, otherwise it is the smallest eigenvalue of the matrix counterpart of the first stage *F*. Critical values based on the Monte Carlo analysis of Stock and Yogo (2003) are shown when available.

The R-squared value printed for models estimated via two-stage least squares is the square of the correlation between the dependent variable and the fitted values.

As alternatives to TSLS, the model may be estimated via Limited Information Maximum Likelihood (the `--liml` option) or via the Generalized Method of Moments (`--gmm` option). Note that if the model is just identified these methods should produce the same results as TSLS, but if it is over-identified the results will differ in general.

If GMM estimation is selected, the following additional options become available:

- `--two-step`: perform two-step GMM rather than the default of one-step.

- `--iterate`: Iterate GMM to convergence.

- `--weights=`*Wmat*: specify a square matrix of weights to be used when computing the GMM criterion function. The dimension of this matrix must equal the number of instruments. The default is an appropriately sized identity matrix.

Menu path: /Model/Other linear models/Two-Stage Least Squares

**var**

| Arguments: | *order ylist* [ ; *xlist* ] |
| --- | --- |
| Options: | `--nc` (do not include a constant) |
| | `--trend` (include a linear trend) |
| | `--seasonals` (include seasonal dummy variables) |
| | `--robust` (robust standard errors) |
| | `--robust-hac` (HAC standard errors) |
| | `--quiet` (skip output of individual equations) |
| | `--silent` (don't print anything) |
| | `--impulse-responses` (print impulse responses) |
| | `--variance-decomp` (print variance decompositions) |
| | `--lagselect` (show information criteria for lag selection) |
| Examples: | `var 4 x1 x2 x3 ; time mydum` |
| | `var 4 x1 x2 x3 --seasonals` |
| | `var 12 x1 x2 x3 --lagselect` |

Sets up and estimates (using OLS) a vector autoregression (VAR). The first argument specifies the lag order — or the maximum lag order in case the `--lagselect` option is given (see below). The order may be given numerically, or as the name of a pre-existing scalar variable. Then follows the setup for the first equation. Don't include lags among the elements of *ylist* — they will be added automatically. The semi-colon separates the stochastic variables, for which *order* lags will be included, from any exogenous variables in *xlist*. Note that a constant is included automatically unless you give the `--nc` flag, a trend can be added with the `--trend` flag, and seasonal dummy variables may be added using the `--seasonals` flag.

A separate regression is reported for each variable in *ylist*. Output for each equation includes *F*-tests for zero restrictions on all lags of each of the variables, an *F*-test for the significance of the maximum lag, and, if the `--impulse-responses` flag is given, forecast variance decompositions and impulse responses.

Forecast variance decompositions and impulse responses are based on the Cholesky decomposition of the contemporaneous covariance matrix, and in this context the order in which the (stochastic) variables are given matters. The first variable in the list is assumed to be "most exogenous" within-period. The horizon for variance decompositions and impulse responses can be set using the set command. For retrieval of a specified impulse response function in matrix form, see the irf function.

If the `--robust` option is given, standard errors are corrected for heteroskedasticity. Alternatively, the `--robust-hac` option can be given to produce standard errors that are robust with respect to both heteroskedasticity and autocorrelation (HAC). In general the latter correction should not be needed if the VAR includes sufficient lags.

If the `--lagselect` option is given, the first parameter to the var command is taken as the maximum lag order. Output consists of a table showing the values of the Akaike (AIC), Schwartz (BIC) and Hannan–Quinn (HQC) information criteria computed from VARs of order 1 to the given maximum. This is intended to help with the selection of the optimal lag order. The usual VAR output is not presented. The table of information criteria may be retrieved as a matrix via the `$test` accessor.

Menu path: /Model/Time series/Vector autoregression

**varlist**

| Options: | `--scalars` (list scalars) |
| --- | --- |
| | `--accessors` (list accessor variables) |

By default, prints a listing of the (series) variables currently available. `list` and `ls` are synonyms.

If the `--scalars` option is given, prints a listing of any currently defined scalar variables and

their values. Otherwise, if the `--accessors` option is given, prints a list of the internal variables currently available via accessors such as $nobs and $uhat.

### vartest

  Arguments:    *var1 var2*

Calculates the *F* statistic for the null hypothesis that the population variances for the variables *var1* and *var2* are equal, and shows its p-value.

Menu path: /Model/Bivariate tests/Difference of variances

### vecm

  Arguments:    *order rank ylist* [ ; *xlist* ] [ ; *rxlist* ]
  Options:      `--nc` (no constant)
                `--rc` (restricted constant)
                `--uc` (unrestricted constant)
                `--crt` (constant and restricted trend)
                `--ct` (constant and unrestricted trend)
                `--seasonals` (include centered seasonal dummies)
                `--quiet` (skip output of individual equations)
                `--silent` (don't print anything)
                `--impulse-responses` (print impulse responses)
                `--variance-decomp` (print variance decompositions)
  Examples:     `vecm 4 1 Y1 Y2 Y3`
                `vecm 3 2 Y1 Y2 Y3 --rc`
                `vecm 3 2 Y1 Y2 Y3 ; X1 --rc`
                See also `denmark.inp`, `hamilton.inp`

A VECM is a form of vector autoregression or VAR (see var), applicable where the variables in the model are individually integrated of order 1 (that is, are random walks, with or without drift), but exhibit cointegration. This command is closely related to the Johansen test for cointegration (see coint2).

The *order* parameter to this command represents the lag order of the VAR system. The number of lags in the VECM itself (where the dependent variable is given as a first difference) is one less than *order*.

The *rank* parameter represents the cointegration rank, or in other words the number of cointegrating vectors. This must be greater than zero and less than or equal to (generally, less than) the number of endogenous variables given in *ylist*.

*ylist* supplies the list of endogenous variables, in levels. The inclusion of deterministic terms in the model is controlled by the option flags. The default if no option is specified is to include an "unrestricted constant", which allows for the presence of a non-zero intercept in the cointegrating relations as well as a trend in the levels of the endogenous variables. In the literature stemming from the work of Johansen (see for example his 1995 book) this is often referred to as "case 3". The first four options given above, which are mutually exclusive, produce cases 1, 2, 4 and 5 respectively. The meaning of these cases and the criteria for selecting a case are explained in the *Gretl User's Guide*.

The optional lists *xlist* and *rxlist* allow you to specify sets of exogenous variables which enter the model either unrestrictedly (*xlist*) or restricted to the cointegration space (*rxlist*). These lists are separated from *ylist* and from each other by semicolons.

The `--seasonals` option, which may be combined with any of the other options, specifies the inclusion of a set of centered seasonal dummy variables. This option is available only for quarterly or monthly data.

The first example above specifies a VECM with lag order 4 and a single cointegrating vector.

The endogenous variables are Y1, Y2 and Y3.  The second example uses the same variables but specifies a lag order of 3 and two cointegrating vectors; it also specifies a "restricted constant", which is appropriate if the cointegrating vectors may have a non-zero intercept but the Y variables have no trend.

Following estimation of a VECM some special accessors are available: $jalpha, $jbeta and $jvbeta retrieve, respectively, the $\alpha$ and $\beta$ matrices and the estimated variance of $\beta$.  For retrieval of a specified impulse response function in matrix form, see the irf function.

Menu path: /Model/Time series/VECM

### vif

Must follow the estimation of a model which includes at least two independent variables. Calculates and displays the Variance Inflation Factors (VIFs) for the regressors. The VIF for regressor $j$ is defined as

$$\frac{1}{1 - R_j^2}$$

where $R_j$ is the coefficient of multiple correlation between regressor $j$ and the other regressors. The factor has a minimum value of 1.0 when the variable in question is orthogonal to the other independent variables. Neter *et al.* (1990) suggest inspecting the largest VIF as a diagnostic for collinearity; a value greater than 10 is sometimes taken as indicating a problematic degree of collinearity.

Menu path: Model window, /Tests/Collinearity

### wls

| Arguments: | *wtvar depvar indepvars* |
|---|---|
| Options: | --vcv (print covariance matrix) |
| | --robust (robust standard errors) |
| | --quiet (suppress printing of results) |

Computes weighted least squares (WLS) estimates using *wtvar* as the weight, *depvar* as the dependent variable, and *indepvars* as the list of independent variables. Let *w* denote the positive square root of wtvar; then WLS is basically equivalent to an OLS regression of *w * depvar* on *w * indepvars*. The *R*-squared, however, is calculated in a special manner, namely as

$$R^2 = 1 - \frac{\text{ESS}}{\text{WTSS}}$$

where ESS is the error sum of squares (sum of squared residuals) from the weighted regression and WTSS denotes the "weighted total sum of squares", which equals the sum of squared residuals from a regression of the weighted dependent variable on the weighted constant alone.

If *wtvar* is a dummy variable, WLS estimation is equivalent to eliminating all observations with value zero for *wtvar*.

Menu path: /Model/Other linear models/Weighted Least Squares

### xcorrgm

| Arguments: | *var1 var2* [ *order* ] |
|---|---|
| Example: | xcorrgm x y 12 |

Prints and graphs the cross-correlogram for variables *var1* and *var2*, which may be specified by name or number. The values are the sample correlation coefficients between the current value of *var1* and successive leads and lags of *var2*.

If an *order* value is specified the length of the cross-correlogram is limited to at most that number of leads and lags, otherwise the length is determined automatically, as a function of the frequency of the data and the number of observations.

Menu path: /View/Cross-correlogram

Other access: Main window pop-up menu (multiple selection)

**xtab**

| | |
|---|---|
| Arguments: | *ylist* [ ; *xlist* ] |
| Options: | `--row` (display row percentages) |
| | `--column` (display column percentages) |
| | `--zeros` (display zero entries) |
| | `--matrix`=*matname* (use frequencies from named matrix) |

Displays a contingency table or cross-tabulation for each combination of the variables included in *ylist*; if a second list *xlist* is given, each variable in *ylist* is cross-tabulated by row against each variable in *xlist* (by column). Variables in these lists can be referenced by name or by number. Note that all the variables must have been marked as discrete. Alternatively, if the `--matrix` option is given, treat the named matrix as a precomputed set of frequencies and display this as a cross-tabulation.

By default the cell entries are given as frequency counts. The `--row` and `--column` options (which are mutually exclusive), replace the counts with the percentages for each row or column, respectively. By default, cells with a zero count are left blank; the `--zeros` option, which has the effect of showing zero counts explicitly, may be useful for importing the table into another program, such as a spreadsheet.

Pearson's chi-square test for independence is displayed if the expected frequency under independence is at least 1.0e-7 for all cells. A common rule of thumb for the validity of this statistic is that at least 80 percent of cells should have expected frequencies of 5 or greater; if this criterion is not met a warning is printed.

If the contingency table is 2 by 2, Fisher's Exact Test for independence is computed. Note that this test is based on the assumption that the row and column totals are fixed, which may or may not be appropriate depending on how the data were generated. The left p-value should be used when the alternative to independence is negative association (values tend to cluster in the lower left and upper right cells); the right p-value should be used if the alternative is positive association. The two-tailed p-value for this test is calculated by method (b) in section 2.1 of Agresti (1992): it is the sum of the probabilities of all possible tables having the given row and column totals and having a probability less than or equal to that of the observed table.

## 1.3   Commands by topic

The following sections show the available commands grouped by topic.

**Estimation**

| | | | |
|---|---|---|---|
| `ar` | Autoregressive estimation | `ar1` | AR(1) estimation |
| `arbond` | Arellano-Bond | `arch` | ARCH model |
| `arima` | ARMA model | `biprobit` | Bivariate probit |
| `dpanel` | Dynamic panel models | `duration` | Duration models |
| `equation` | Define equation within a system | `estimate` | Estimate system of equations |
| `garch` | GARCH model | `gmm` | GMM estimation |
| `heckit` | Heckman selection model | `hsk` | Heteroskedasticity-corrected estimates |
| `intreg` | Interval regression model | `kalman` | Kalman filter |
| `lad` | Least Absolute Deviation estimation | `logistic` | Logistic regression |
| `logit` | Logit regression | `mle` | Maximum likelihood estimation |
| `mpols` | Multiple-precision OLS | `negbin` | Negative Binomial regression |
| `nls` | Nonlinear Least Squares | `ols` | Ordinary Least Squares |

| | | | |
|---|---|---|---|
| panel | Panel models | poisson | Poisson estimation |
| probit | Probit model | quantreg | Quantile regression |
| system | Systems of equations | tobit | Tobit model |
| tsls | Instrumental variables regression | var | Vector Autoregression |
| vecm | Vector Error Correction Model | wls | Weighted Least Squares |

## Tests

| | | | |
|---|---|---|---|
| add | Add variables to model | adf | Augmented Dickey-Fuller test |
| chow | Chow test | coeffsum | Sum of coefficients |
| coint | Engle-Granger cointegration test | coint2 | Johansen cointegration test |
| cusum | CUSUM test | difftest | Nonparametric tests for differences |
| hausman | Panel diagnostics | kpss | KPSS stationarity test |
| leverage | Influential observations | levinlin | Levin-Lin-Chu test |
| meantest | Difference of means | modtest | Model tests |
| normtest | Normality test | omit | Omit variables |
| qlrtest | Quandt likelihood ratio test | reset | Ramsey's RESET |
| restrict | Testing restrictions | runs | Runs test |
| vartest | Difference of variances | vif | Variance Inflation Factors |

## Transformations

| | | | |
|---|---|---|---|
| diff | First differences | discrete | Mark variables as discrete |
| dummify | Create sets of dummies | lags | Create lags |
| ldiff | Log-differences | logs | Create logs |
| orthdev | Orthogonal deviations | sdiff | Seasonal differencing |
| square | Create squares of variables | | |

## Statistics

| | | | |
|---|---|---|---|
| anova | ANOVA | corr | Correlation coefficients |
| corrgm | Correlogram | fractint | Fractional integration |
| freq | Frequency distribution | hurst | Hurst exponent |
| mahal | Mahalanobis distances | pca | Principal Components Analysis |
| pergm | Periodogram | spearman | Spearmans's rank correlation |
| summary | Descriptive statistics | xcorrgm | Cross-correlogram |
| xtab | Cross-tabulate variables | | |

## Dataset

| | | | |
|---|---|---|---|
| append | Append data | data | Import from database |
| dataset | Manipulate the dataset | delete | Delete variables |
| genr | Generate a new variable | info | Information on data set |
| labels | Print labels for variables | nulldata | Creating a blank dataset |
| open | Open a data file | rename | Rename variables |
| setinfo | Edit attributes of variable | setobs | Set frequency and starting observation |
| setmiss | Missing value code | smpl | Set the sample range |
| store | Save data | varlist | Listing of variables |

## Graphs

| | | | |
|---|---|---|---|
| boxplot | Boxplots | gnuplot | Create a gnuplot graph |
| graphpg | Gretl graph page | qqplot | Q-Q plot |
| rmplot | Range-mean plot | scatters | Multiple pairwise graphs |
| textplot | ASCII plot | | |

## Printing

| | | | |
|---|---|---|---|
| eqnprint | Print model as equation | modprint | Print a user-defined model |
| outfile | Direct printing to file | print | Print data or strings |
| printf | Formatted printing | sprintf | Printing to a string |
| tabprint | Print model in tabular form | | |

## Prediction

fcast    Generate forecasts

## Programming

| | | | |
|---|---|---|---|
| break | Break from loop | catch | Catch errors |
| clear | | debug | Debugging |
| elif | Flow control | else | |
| end | End block of commands | endif | Flow control |
| endloop | End a command loop | foreign | Non-native script |
| function | Define a function | if | Flow control |
| include | Include function definitions | loop | Start a command loop |
| makepkg | Make function package | run | Execute a script |
| set | Set program parameters | sscanf | Scanning a string |

## Utilities

| | | | |
|---|---|---|---|
| help | Help on commands | modeltab | The model table |
| pvalue | Compute p-values | quit | Exit the program |
| shell | Execute shell commands | | |

# Chapter 2

# Gretl functions

## 2.1 Introduction

This chapter presents two alphabetical listings: first, the "accessors" which enable the user to retrieve the values of internal variables; and second, the functions proper that are available in the context of the `genr` command.

## 2.2 Accessors

**$ahat**

  Output:    series

Must follow the estimation of a fixed-effect panel data model. Returns a series containing the estimates of the individual fixed effects (per-unit intercepts).

**$aic**

  Output:    scalar

Returns the Akaike Information Criterion for the last estimated model, if available. See the *Gretl User's Guide* for details of the calculation.

**$bic**

  Output:    scalar

Returns Schwarz's Bayesian Information Criterion for the last estimated model, if available. See the *Gretl User's Guide* for details of the calculation.

**$chisq**

  Output:    scalar

Returns the overall chi-square statistic from the last estimated model, if available.

**$coeff**

  Output:      matrix or scalar
  Argument:    *s* (name of coefficient, optional)

With no arguments, `$coeff` returns a column vector containing the estimated coefficients for the last model. With the optional string argument it returns a scalar, namely the estimated parameter named *s*. See also $stderr, $vcv.

Example:

```
open bjg
arima 0 1 1 ; 0 1 1 ; lg
b = $coeff              # gets a vector
macoef = $coeff(theta_1) # gets a scalar
```

If the "model" in question is actually a system, the result depends on the characteristics of the system: for VARs and VECMs the value returned is a matrix with one column per equation, otherwise it is a column vector containing the coefficients from the first equation followed by those from the second equation, and so on.

### $command

Output:   string

Must follow the estimation of a model; returns the command word, for example `ols` or `probit`.

### $compan

Output:   matrix

Must follow the estimation of a VAR or a VECM; returns the companion matrix.

### $datatype

Output:   scalar

Returns an integer value representing the sort of dataset that is currently loaded: 0 = no data; 1 = cross-sectional (undated) data; 2 = time-series data; 3 = panel data.

### $depvar

Output:   string

Must follow the estimation of a single-equation model; returns the name of the dependent variable.

### $df

Output:   scalar

Returns the degrees of freedom of the last estimated model. If the last model was in fact a system of equations, the value returned is the degrees of freedom per equation; if this differs across the equations then the value given is the number of observations minus the mean number of coefficients per equation (rounded up to the nearest integer).

### $dwpval

Output:   scalar

Returns the p-value for the Durbin–Watson statistic for the model last estimated, if available. This is computed using the Imhof procedure.

### $ec

Output:   matrix

Must follow the estimation of a VECM; returns a matrix containing the error correction terms. The number of rows equals the number of observations used and the number of columns equals the cointegration rank of the system.

### $error

Output:   scalar

Returns the program's internal error code, which will be non-zero in case an error has occurred but has been trapped using catch. Note that using this accessor causes the internal error code to be reset to zero. See also errmsg. If you want to get the error message associated with a given `$error` you need to store the value in a temporary variable, as in

```
        errval = $error
        if (errval)
          printf "Got error %d (%s)\n", errval, errmsg(errval);
        endif
```

**$ess**

  Output:   scalar

Returns the error sum of squares of the last estimated model, if available.

**$evals**

  Output:   matrix

Must follow the estimation of a VECM; returns a vector containing the eigenvalues that are used in computing the trace test for cointegration.

**$fcast**

  Output:   matrix

Must follow the fcast forecasting command; returns the forecast values as a matrix. If the model on which the forecast was based is a system of equations the returned matrix will have one column per equation, otherwise it is a column vector.

**$fcerr**

  Output:   matrix

Must follow the fcast forecasting command; returns the standard errors of the forecasts, if available, as a matrix. If the model on which the forecast was based is a system of equations the returned matrix will have one column per equation, otherwise it is a column vector.

**$fevd**

  Output:   matrix

Must follow estimation of a VAR. Returns a matrix containing the forecast error variance decomposition. This matrix has $h$ rows where $h$ is the forecast horizon, which can be chosen using set horizon or otherwise is set automatically based on the frequency of the data. For a VAR with $p$ variables, the matrix has $p^2$ columns. The fraction of the forecast error for variable $i$ attributable to innovation in variable $j$ is found in column $(i - 1)p + j$.

**$Fstat**

  Output:   scalar

Returns the overall F-statistic from the last estimated model, if available.

**$gmmcrit**

  Output:   scalar

Must follow a gmm block. Returns the value of the objective function at its minimum.

**$h**

  Output:   series

Must follow a garch command. Returns the estimated conditional variance series.

**$hausman**

   Output:   row vector

Must follow estimation of a model via either `tsls` or `panel` with the random effects option. Returns a $1 \times 3$ vector containing the value of the Hausman test statistic, the corresponding degrees of freedom and the p-value for the test, in that order.

**$hqc**

   Output:   scalar

Returns the Hannan-Quinn Information Criterion for the last estimated model, if available. See the *Gretl User's Guide* for details of the calculation.

**$jalpha**

   Output:   matrix

Must follow the estimation of a VECM, and returns the loadings matrix. It has as many rows as variables in the VECM and as many columns as the cointegration rank.

**$jbeta**

   Output:   matrix

Must follow the estimation of a VECM, and returns the cointegration matrix. It has as many rows as variables in the VECM (plus the number of exogenous variables that are restricted to the cointegration space, if any), and as many columns as the cointegration rank.

**$jvbeta**

   Output:   square matrix

Must follow the estimation of a VECM, and returns the estimated covariance matrix for the elements of the cointegration vectors.

In the case of unrestricted estimation, this matrix has a number of rows equal to the unrestricted elements of the cointegration space after the Phillips normalization. If, however, a restricted system is estimated via the `restrict` command with the `--full` option, a singular matrix with $(n + m)r$ rows will be returned ($n$ being the number of endogenous variables, $m$ the number of exogenous variables that are restricted to the cointegration space, and $r$ the cointegration rank).

Example: the code

```
open denmark.gdt
vecm 2 1 LRM LRY IBO IDE --rc --seasonals -q
s0 = $jvbeta

restrict --full
b[1,1] = 1
b[1,2] = -1
b[1,3] + b[1,4] = 0
end restrict
s1 = $jvbeta

print s0
print s1
```

produces the following output.

```
s0 (4 x 4)

    0.019751      0.029816    -0.00044837     -0.12227
    0.029816       0.31005     -0.45823       -0.18526
 -0.00044837      -0.45823       1.2169       -0.035437
   -0.12227       -0.18526     -0.035437        0.76062

s1 (5 x 5)

    0.0000        0.0000        0.0000         0.0000        0.0000
    0.0000        0.0000        0.0000         0.0000        0.0000
    0.0000        0.0000        0.27398       -0.27398      -0.019059
    0.0000        0.0000       -0.27398        0.27398       0.019059
    0.0000        0.0000       -0.019059       0.019059      0.0014180
```

**$llt**

Output:    series

For selected models estimated via Maximum Likelihood, returns the series of per-observation log-likelihood values. At present this is supported only for binary logit and probit, tobit and heckit.

**$lnl**

Output:    scalar

Returns the log-likelihood for the last estimated model (where applicable).

**$mnlprobs**

Output:    matrix

Following estimation of a multinomial logit model (only), retrieves a matrix holding the estimated probabilities of each possible outcome at each observation in the model's sample range. Each row represents an observation and each column an outcome.

**$ncoeff**

Output:    scalar

Returns the total number of coefficients estimated in the last model.

**$nobs**

Output:    scalar

Returns the number of observations in the currently selected sample.

**$nvars**

Output:    scalar

Returns the number of variables in the dataset (including the constant).

**$pd**

Output:    scalar

Returns the frequency or periodicity of the data (e.g. 4 for quarterly data). In the case of panel data the value returned is the time-series length.

**$pvalue**

   Output:   scalar or matrix

Returns the p-value of the test statistic that was generated by the last explicit hypothesis-testing command, if any (e.g. chow). See the *Gretl User's Guide* for details.

In most cases the return value is a scalar but sometimes it is a matrix (for example, the trace and lambda-max p-values from the Johansen cointegration test); in that case the values in the matrix are laid out in the same pattern as the printed results.

See also $test.

**$rho**

   Output:     scalar
   Argument:   $n$ (scalar, optional)

Without arguments, returns the first-order autoregressive coefficient for the residuals of the last model. After estimating a model via the `ar` command, the syntax `$rho(n)` returns the corresponding estimate of $\rho(n)$.

**$rsq**

   Output:   scalar

Returns the unadjusted $R^2$ from the last estimated model, if available.

**$sample**

   Output:   series

Must follow estimation of a single-equation model. Returns a dummy series with value 1 for observations used in estimation, 0 for observations within the currently defined sample range but not used (presumably because of missing values), and NA for observations outside of the current range.

If you wish to compute statistics based on the sample that was used for a given model, you can do, for example:

```
ols y 0 xlist
genr sdum = $sample
smpl sdum --dummy
```

**$sargan**

   Output:   row vector

Must follow a `tsls` command. Returns a $1 \times 3$ vector, containing the value of the Sargan over-identification test statistic, the corresponding degrees of freedom and p-value, in that order.

**$sigma**

   Output:   scalar or matrix

Requires that a model has been estimated. If the last model was a single equation, returns the (scalar) Standard Error of the Regression (or in other words, the standard deviation of the residuals, with an appropriate degrees of freedom correction). If the last model was a system of equations, returns the cross-equation covariance matrix of the residuals.

**$stderr**

Output: matrix or scalar

Argument: *s* (name of coefficient, optional)

With no arguments, `$stderr` returns a column vector containing the standard error of the coefficients for the last model. With the optional string argument it returns a scalar, namely the standard error of the parameter named *s*.

If the "model" in question is actually a system, the result depends on the characteristics of the system: for VARs and VECMs the value returned is a matrix with one column per equation, otherwise it is a column vector containing the coefficients from the first equation followed by those from the second equation, and so on.

See also $coeff, $vcv.

**$stopwatch**

Output: scalar

Must be preceded by `set stopwatch`, which activates the measurement of CPU time. The first use of this accessor yields the seconds of CPU time that have elapsed since the `set stopwatch` command. At each access the clock is reset, so subsequent uses of `$stopwatch` yield the seconds of CPU time since the previous access.

**$sysA**

Output: matrix

Must follow estimation of a simultaneous equations system. Returns the matrix of coefficients on the lagged endogenous variables, if any, in the structural form of the system. See the system command.

**$sysB**

Output: matrix

Must follow estimation of a simultaneous equations system. Returns the matrix of coefficients on the exogenous variables in the structural form of the system. See the system command.

**$sysGamma**

Output: matrix

Must follow estimation of a simultaneous equations system. Returns the matrix of coefficients on the contemporaneous endogenous variables in the structural form of the system. See the system command.

**$T**

Output: scalar

Returns the number of observations used in estimating the last model.

**$t1**

Output: scalar

Returns the 1-based index of the first observation in the currently selected sample.

**$t2**

Output: scalar

Returns the 1-based index of the last observation in the currently selected sample.

**$test**

  Output:    scalar or matrix

Returns the value of the test statistic that was generated by the last explicit hypothesis-testing command, if any (e.g. chow). See the *Gretl User's Guide* for details.

In most cases the return value is a scalar but sometimes it is a matrix (for example, the trace and lambda-max statistics from the Johansen cointegration test); in that case the values in the matrix are laid out in the same pattern as the printed results.

See also $pvalue.

**$trsq**

  Output:    scalar

Returns $TR^2$ (sample size times R-squared) from the last model, if available.

**$uhat**

  Output:    series

Returns the residuals from the last model. This may have different meanings for different estimators. For example, after an ARMA estimation $uhat will contain the one-step-ahead forecast error; after a probit model, it will contain the generalized residuals.

If the "model" in question is actually a system (a VAR or VECM, or system of simultaneous equations), $uhat with no parameters retrieves the matrix of residuals, one column per equation.

**$unit**

  Output:    series

Valid for panel datasets only. Returns a series with value 1 for all observations on the first unit or group, 2 for observations on the second unit, and so on.

**$vcv**

  Output:        matrix or scalar
  Arguments:   *s1* (name of coefficient, optional)
              *s2* (name of coefficient, optional)

With no arguments, $vcv returns a square matrix containing the estimated covariance matrix for the coefficients of the last model. If the last model was a single equation, then you may supply the names of two parameters in parentheses to retrieve the estimated covariance between the parameters named *s1* and *s2*. See also $coeff, $stderr.

This accessor is not available for VARs or VECMs; for models of that sort see $sigma and $xtxinv.

**$vecGamma**

  Output:    matrix

Must follow the estimation of a VECM; returns a matrix in which the Gamma matrices (coefficients on the lagged differences of the cointegrated variables) are stacked side by side. Each row represents an equation; for a VECM of lag order $p$ there are $p - 1$ sub-matrices.

**$version**

  Output:    scalar

Returns an integer value that codes for the program version. The gretl version string takes the form x.y.z (for example, 1.7.6). The return value from this accessor is formed as 10000*x + 100*y + z, so that 1.7.6 translates as 10706.

### $vma

   Output:   matrix

Must follow the estimation of a VAR or a VECM; returns a matrix containing the VMA representation up to the order specified via the `set horizon` command. See the *Gretl User's Guide* for details.

### $windows

   Output:   scalar

Returns 1 if gretl is running on MS Windows, otherwise 0. By conditioning on the value of this variable you can write shell calls that are portable across different operating systems.

Also see the shell command.

### $xlist

   Output:   list

Returns the list of regressors from the last model (for single-equation models only).

### $xtxinv

   Output:   matrix

Following estimation of a VAR or VECM (only), returns $X'X^{-1}$, where $X$ is the common matrix of regressors used in each of the equations. This accessor is not available for a VECM estimated with a restriction imposed on $\alpha$, the "loadings" matrix.

### $yhat

   Output:   series

Returns the fitted values from the last regression.

## 2.3 Functions proper

### abs

   Output:      same type as input
   Argument:   *x* (scalar, series or matrix)

Returns the absolute value of *x*.

### acos

   Output:      same type as input
   Argument:   *x* (scalar, series or matrix)

Returns the arc cosine of *x*, that is, the value whose cosine is *x*. The result is in radians; the input should be in the range $-1$ to 1.

**acosh**

  Output:      same type as input
  Argument:   $x$ (scalar, series or matrix)

Returns the inverse hyperbolic cosine of $x$ (positive solution). $x$ should be greater than 1; otherwise, NA is returned. See also cosh.

**argname**

  Output:      string
  Argument:   $s$ (string)

For $s$ the name of a parameter to a user-defined function, returns the name of the corresponding argument, or an empty string if the argument was anonymous.

**asin**

  Output:      same type as input
  Argument:   $x$ (scalar, series or matrix)

Returns the arc sine of $x$, that is, the value whose sine is $x$. The result is in radians; the input should be in the range $-1$ to 1.

**asinh**

  Output:      same type as input
  Argument:   $x$ (scalar, series or matrix)

Returns the inverse hyperbolic sine of $x$. See also sinh.

**atan**

  Output:      same type as input
  Argument:   $x$ (scalar, series or matrix)

Returns the arc tangent of $x$, that is, the value whose tangent is $x$. The result is in radians.

**atanh**

  Output:      same type as input
  Argument:   $x$ (scalar, series or matrix)

Returns the inverse hyperbolic tangent of $x$. See also tanh.

**bessel**

  Output:      same type as input
  Arguments:   *type* (character)
              $v$ (scalar)
              $x$ (scalar, series or matrix)

Computes one of the Bessel function variants for order $v$ and argument $x$. The return value is of the same type as $x$. The specific function is selected by the first argument, which must be J, Y, I, or K. A good discussion of the Bessel functions can be found on Wikipedia; here we give a brief account.

case J: Bessel function of the first kind. Resembles a damped sine wave. Defined for real $v$ and $x$, but if $x$ is negative then $v$ must be an integer.

case Y: Bessel function of the second kind. Defined for real $v$ and $x$ but has a singularity at $x = 0$.

case I: Modified Bessel function of the first kind. An exponentially growing function. Acceptable arguments are as for case J.

case K: Modified Bessel function of the second kind. An exponentially decaying function. Diverges at $x = 0$ and is not defined for negative $x$. Symmetric around $v = 0$.

### BFGSmax

| | |
|---|---|
| Output: | scalar |
| Arguments: | $b$ (vector) |
| | $f$ (function call) |
| | $g$ (function call, optional) |

Numerical maximization via the method of Broyden, Fletcher, Goldfarb and Shanno. The vector $b$ should hold the initial values of a set of parameters, and the argument $f$ should specify a call to a function that calculates the (scalar) criterion to be maximized, given the current parameter values and any other relevant data. If the object is in fact minimization, this function should return the negative of the criterion. On successful completion, BFGSmax returns the maximized value of the criterion, and $b$ holds the parameter values which produce the maximum.

The optional third argument provides a means of supplying analytical derivatives (otherwise the gradient is computed numerically). The gradient function call $g$ must have as its first argument a pre-defined matrix that is of the correct size to contain the gradient, given in pointer form. It also must take the parameter vector as an argument (in pointer form or otherwise). Other arguments are optional.

For more details and examples see the chapter on numerical methods in the *Gretl User's Guide*. See also NRmax, fdjac, simann.

### bkfilt

| | |
|---|---|
| Output: | series |
| Arguments: | $y$ (series) |
| | $f1$ (scalar, optional) |
| | $f2$ (scalar, optional) |
| | $k$ (scalar, optional) |

Returns the result from application of the Baxter–King bandpass filter to the series $y$. The optional parameters $f1$ and $f2$ represent, respectively, the lower and upper bounds of the range of frequencies to extract, while $k$ is the approximation order to be used. If these arguments are not supplied then the following default values are used: $f1 = 8$, $f1 = 32$, $k = 8$. See also hpfilt.

### boxcox

| | |
|---|---|
| Output: | series |
| Arguments: | $y$ (series) |
| | $d$ (scalar) |

Returns the Box–Cox transformation with parameter $d$ for the positive series $y$.

$$y_t^{(d)} = \begin{cases} \frac{y_t^d - 1}{d} & \text{if} \quad d \neq 0 \\ \log(y_t) & \text{if} \quad d = 0 \end{cases}$$

### bwfilt

| | |
|---|---|
| Output: | series |
| Arguments: | $y$ (series) |
| | $n$ (scalar) |
| | $omega$ (scalar) |

Returns the result from application of a low-pass Butterworth filter with order $n$ and frequency cutoff *omega* to the series $y$. The cutoff is expressed in degrees and must be greater than 0 and less than 180. Smaller cutoff values restrict the pass-band to lower frequencies and hence produce a smoother trend. Higher values of $n$ produce a sharper cutoff, at the cost of possible numerical instability.

Inspecting the periodogram of the target series is a useful preliminary when you wish to apply this function. See the *Gretl User's Guide* for details. See also bkfilt, hpfilt.

### cdemean

| | |
|---|---|
| Output: | matrix |
| Argument: | $X$ (matrix) |

Centers the columns of matrix $X$ around their means.

### cdf

| | |
|---|---|
| Output: | same type as input |
| Arguments: | $c$ (character) |
| | ... (see below) |
| | $x$ (scalar, series or matrix) |
| Examples: | p1 = cdf(N, –2.5) |
| | p2 = cdf(X, 3, 5.67) |
| | p3 = cdf(D, 0.25, –1, 1) |

Cumulative distribution function calculator. Returns $P(X \leq x)$, where the distribution $X$ is determined by the character $c$. Between the arguments $c$ and $x$, zero or more additional scalar arguments are required to specify the parameters of the distribution, as follows.

| *Distribution* | *c* | *Arg* 2 | *Arg* 3 |
|---|---|---|---|
| Standard normal | z, n or N | – | – |
| Bivariate normal | D | $\rho$ | – |
| Student's $t$ (central) | t | degrees of freedom | – |
| Chi square | c, x or X | degrees of freedom | – |
| Snedecor's $F$ | f or F | df (num.) | df (den.) |
| Gamma | g or G | shape | scale |
| Binomial | b or B | probability | trials |
| Poisson | p or P | mean | – |
| Weibull | w or W | shape | scale |
| Generalized Error | E | shape | – |

Note that most cases have aliases to help memorizing the codes. The bivariate normal case is special: the syntax is x = cdf(D, rho, z1, z2) where rho is the correlation between the variables z1 and z2.

The parametrization gretl uses for the Gamma random variate implies that its density function can be written as

$$f(x; k, \theta) = \frac{x^{k-1}}{\theta^k} \frac{e^{-x/\theta}}{\Gamma(k)}$$

where $k > 0$ is the shape parameter and $\theta > 0$ is the scale parameter.

See also pdf, critical, invcdf, pvalue.

**cdiv**

Output: matrix
Arguments: *X* (matrix)
*Y* (matrix)

Complex division. The two arguments must have the same number of rows, $n$, and either one or two columns. The first column contains the real part and the second (if present) the imaginary part. The return value is an $n \times 2$ matrix or, if the result has no imaginary part, an $n$-vector. See also cmult.

**ceil**

Output: same type as input
Argument: *x* (scalar, series or matrix)

Ceiling function: returns the smallest integer greater than or equal to *x*. See also floor, int.

**cholesky**

Output: square matrix
Argument: *A* (symmetric matrix)

Peforms a Cholesky decomposition of the matrix *A*, which is assumed to be symmetric and positive definite. The result is a lower-triangular matrix *L* which satisfies $A = LL'$. The function will fail if *A* is not symmetric or not positive definite. See also psdroot.

**chowlin**

Output: matrix
Arguments: *Y* (matrix)
*xfac* (scalar)
*X* (matrix, optional)

Expands the input data, *Y*, to a higher frequency, using the interpolation method of Chow and Lin (1971). It is assumed that the columns of *Y* represent data series; the returned matrix has as many columns as *Y* and *xfac* times as many rows.

The second argument represents the expansion factor: it should be 3 for expansion from quarterly to monthly or 4 for expansion from annual to quarterly, these being the only supported factors. The optional third argument may be used to provide a matrix of covariates at the higher (target) frequency.

The regressors used by default are a constant and quadratic trend. If *X* is provided, its columns are used as additional regressors; it is an error if the number of rows in *X* does not equal *xfac* times the number of rows in *Y*.

**cmult**

Output: matrix
Arguments: *X* (matrix)
*Y* (matrix)

Complex multiplication. The two arguments must have the same number of rows, $n$, and either one or two columns. The first column contains the real part and the second (if present) the imaginary part. The return value is an $n \times 2$ matrix, or, if the result has no imaginary part, an $n$-vector. See also cdiv.

**cnorm**

  Output:     same type as input
  Argument:   $x$ (scalar, series or matrix)

Returns the cumulative distribution function for a standard normal. See also dnorm, qnorm.

**colname**

  Output:     string
  Arguments:  $M$ (matrix)
             $col$ (scalar)

Retrieves the name for column $col$ of matrix $M$. If $M$ has no column names attached the value returned is an empty string; if $col$ is out of bounds for the given matrix an error is flagged. See also colnames.

**colnames**

  Output:     scalar
  Arguments:  $M$ (matrix)
             $s$ (named list or string)

Attaches names to the columns of the $T \times k$ matrix $M$. If $s$ is a named list, the column names are copied from the names of the variables; the list must have $k$ members. If $s$ is a string, it should contain $k$ space-separated sub-strings. The return value is 0 on successful completion, non-zero on error. See also rownames.

**cols**

  Output:     scalar
  Argument:   $X$ (matrix)

The number of columns of $X$. See also mshape, rows, unvech, vec, vech.

**corr**

  Output:     scalar
  Arguments:  $y1$ (series or vector)
             $y2$ (series or vector)

Computes the correlation coefficient between $y1$ and $y2$. The arguments should be either two series, or two vectors of the same length. See also cov, mcov, mcorr.

**corrgm**

  Output:     matrix
  Arguments:  $x$ (series, matrix or list)
             $p$ (scalar)
             $y$ (series or vector, optional)

If only the first two arguments are given, computes the correlogram for $x$ for lags 1 to $p$. Let $k$ represent the number of elements in $x$ (1 if $x$ is a series, the number of columns if $x$ is a matrix, or the number of list-members is $x$ is a list). The return value is a matrix with $p$ rows and $2k$ columns, the first $k$ columns holding the respective autocorrelations and the remainder the respective partial autocorrelations.

If a third argument is given, this function computes the cross-correlogram for each of the $k$ elements in $x$ and $y$, from lead $p$ to lag $p$. The returned matrix has $2p + 1$ rows and $k$ columns. If $x$ is series or list and $y$ is a vector, the vector must have just as many rows as there are observations in the current sample range.

## cos

Output:       same type as input
Argument:     $x$ (scalar, series or matrix)

Returns the cosine of $x$.

## cosh

Output:       same type as input
Argument:     $x$ (scalar, series or matrix)

Returns the hyperbolic cosine of $x$.

$$\cosh x = \frac{e^x + e^{-x}}{2}$$

See also acosh, sinh, tanh.

## cov

Output:       scalar
Arguments:    $y1$ (series or vector)
              $y2$ (series or vector)

Returns the covariance between $y1$ and $y2$. The arguments should be either two series, or two vectors of the same length. See also corr, mcov, mcorr.

## critical

Output:       same type as input
Arguments:    $c$ (character)
              ... (see below)
              $p$ (scalar, series or matrix)
Examples:     c1 = critical(t, 20, 0.025)
              c2 = critical(F, 4, 48, 0.05)

Critical value calculator. Returns $x$ such that $P(X > x) = p$, where the distribution $X$ is determined by the character $c$. Between the arguments $c$ and $p$, zero or more additional scalar arguments are required to specify the parameters of the distribution, as follows.

| Distribution | $c$ | Arg 2 | Arg 3 |
|---|---|---|---|
| Standard normal | z, n or N | – | – |
| Student's $t$ (central) | t | degrees of freedom | – |
| Chi square | c, x or X | degrees of freedom | – |
| Snedecor's $F$ | f or F | df (num.) | df (den.) |
| Binomial | b or B | $p$ | $n$ |
| Poisson | p or P | $\lambda$ | – |

See also cdf, invcdf, pvalue.

## cum

Output:       same type as input
Argument:     $x$ (series or matrix)

Cumulates $x$. When $x$ is a series, produces a series $y_t = \sum_{s=m}^{t} x_s$; the starting point of the summation, $m$, is the first non-missing observation of the currently selected sample. If any

missing values are encountered in $x$, subsequent values of $y$ will be set to missing. When $x$ is a matrix, its elements are cumulated by columns.

See also diff.

**deseas**

| | |
|---|---|
| Output: | series |
| Arguments: | $x$ (series) |
| | $c$ (character, optional) |

Depends on having TRAMO/SEATS or X-12-ARIMA installed. Returns a deseasonalized (seasonally adjusted) version of the input series $x$, which must be a quarterly or monthly time series. To use X-12-ARIMA give X as the second argument; to use TRAMO give T. If the second argument is omitted then X-12-ARIMA is used.

Note that if the input series has no detectable seasonal component this function will fail. Also note that both TRAMO/SEATS and X-12-ARIMA offer numerous options; `deseas` calls them with all options at their default settings. For both programs, the seasonal factors are calculated on the basis of an automatically selected ARIMA model. One difference between the programs which can sometimes make a substantial difference to the results is that by default TRAMO performs a prior adjustment for outliers while X-12-ARIMA does not.

**det**

| | |
|---|---|
| Output: | scalar |
| Argument: | $A$ (square matrix) |

Returns the determinant of $A$, computed via the LU factorization. See also ldet, rcond.

**diag**

| | |
|---|---|
| Output: | matrix |
| Argument: | $X$ (matrix) |

Returns the principal diagonal of $X$ in a column vector. Note: if $X$ is an $m \times n$ marix, the number of elements of the output vector is $\min(m, n)$. See also tr.

**diagcat**

| | |
|---|---|
| Output: | matrix |
| Arguments: | $A$ (matrix) |
| | $B$ (matrix) |

Returns the direct sum of $A$ and $B$, that is a block-diagonal matrix holding $A$ in its north-west corner and $B$ in its south-east corner.

**diff**

| | |
|---|---|
| Output: | same type as input |
| Argument: | $y$ (series, matrix or list) |

Computes first differences. If $y$ is a series, or a list of series, starting values are set to NA. If $y$ is a matrix, differencing is done by columns and starting values are set to 0.

When a list is returned, the individual variables are automatically named according to the template d_*varname* where *varname* is the name of the original series. The name is truncated if necessary, and may be adjusted in case of non-uniqueness in the set of names thus constructed.

See also cum, ldiff, sdiff.

### digamma

Output:     same type as input
Argument:   $x$ (scalar, series or matrix)

Returns the digamma (or Psi) function of $x$, that is $\frac{d\Gamma(x)}{dx}$.

### dnorm

Output:     same type as input
Argument:   $x$ (scalar, series or matrix)

Returns the density of the standard normal distribution at $x$. To get the density for a non-standard normal distribution at $x$, pass the $z$-score of $x$ to the dnorm function and multiply the result by the Jacobian of the $z$ transformation, namely 1 over $\sigma$, as illustrated below:

```
mu = 100
sigma = 5
x = 109
fx = (1/sigma) * dnorm((x-mu)/sigma)
```

See also cnorm, qnorm.

### dsort

Output:     same type as input
Argument:   $x$ (series or vector)

Sorts $x$ in descending order, skipping observations with missing values when $x$ is a series. See also sort, values.

### dummify

Output:     list
Arguments:  $x$ (series)
            *omitval* (scalar, optional)

The argument $x$ should be a discrete series. This function creates a set of dummy variables coding for the distinct values in the series. By default the smallest value is taken as the omitted category and is not explicitly represented.

The optional second argument represents the value of $x$ which should be treated as the omitted category. The effect when a single argument is given is equivalent to dummify(x, min(x)). To produce a full set of dummies, with no omitted category, use dummify(x, NA).

The generated variables are automatically named according to the template D*varname_i* where *varname* is the name of the original series and $i$ is a 1-based index. The original portion of the name is truncated if necessary, and may be adjusted in case of non-uniqueness in the set of names thus constructed.

### eigengen

Output:     matrix
Arguments:  $A$ (square matrix)
            *&U* (reference to matrix, or null)

Computes the eigenvalues, and optionally the right eigenvectors, of the $n \times n$ matrix $A$. If all the eigenvalues are real an $n \times 1$ matrix is returned; otherwise the result is an $n \times 2$ matrix, the first column holding the real components and the second column the imaginary components.

The second argument must be either the name of an existing matrix preceded by & (to indicate the "address" of the matrix in question), in which case an auxiliary result is written to that matrix, or the keyword null, in which case the auxiliary result is not produced.

If a non-null second argument is given, the specified matrix will be over-written with the auxiliary result. (It is not required that the existing matrix be of the right dimensions to receive the result.) It will be organized as follows:

- If the $i$-th eigenvalue is real, the $i$-th column of $U$ will contain the corresponding eigenvector;

- If the $i$-th eigenvalue is complex, the $i$-th column of $U$ will contain the real part of the corresponding eigenvector and the next column the imaginary part. The eigenvector for the conjugate eigenvalue is the conjugate of the eigenvector.

In other words, the eigenvectors are stored in the same order as the eigenvalues, but the real eigenvectors occupy one column, whereas complex eigenvectors take two (the real part comes first); the total number of columns is still $n$, because the conjugate eigenvector is skipped.

See also eigensym, eigsolve, qrdecomp, svd.

### eigensym

  Output:        matrix
  Arguments:    $A$ (symmetric matrix)
                 $\&U$ (reference to matrix, or null)

Works just as eigengen, but the argument $A$ must be symmetric (in which case the calculations can be reduced). The eigenvalues are returned in ascending order.

### eigsolve

  Output:        matrix
  Arguments:    $A$ (symmetric matrix)
                 $B$ (symmetric matrix)
                 $\&U$ (reference to matrix, or null)

Solves the generalized eigenvalue problem $|A - \lambda B| = 0$, where both $A$ and $B$ are symmetric and $B$ is positive definite. The eigenvalues are returned directly, arranged in ascending order. If the optional third argument is given it should be the name of an existing matrix preceded by &; in that case the generalized eigenvectors are written to the named matrix.

### epochday

  Output:        scalar
  Arguments:    *year* (scalar)
                 *month* (scalar)
                 *day* (scalar)

Returns the number of the day in the current epoch specified by year, month and day (which equals 1 for the first of January in the year 1 AD).

### errmsg

  Output:        string
  Argument:     *errno* (scalar)

Retrieves the gretl error message associated with *errno*. See also $error.

**exp**

Output:      same type as input
Argument:    $x$ (scalar, series or matrix)

Returns $e^x$. Note that in case of matrices the function acts element by element. For the matrix exponential function, see mexp.

**fcstats**

Output:      matrix
Arguments:   $y$ (series or vector)
             $f$ (series or vector)

Produces a column vector holding several statistics which may be used for evaluating the series $f$ as a forecast of the series $y$ over the current sample range. Two vectors of the same length may be given in place of two series arguments.

The layout of the returned vector is as follows:

```
1  Mean Error (ME)
2  Mean Squared Error (MSE)
3  Mean Absolute Error (MAE)
4  Mean Percentage Error (MPE)
5  Mean Absolute Percentage Error (MAPE)
6  Theil's U
7  Bias proportion, UM
8  Regression proportion, UR
9  Disturbance proportion, UD
```

For details on the calculation of these statistics, and the interpretation of the $U$ values, please see the *Gretl User's Guide*.

**fdjac**

Output:      matrix
Arguments:   $b$ (column vector)
             $f$ (function call)

Calculates the (forward-difference approximation to the) Jacobian associated with the vector $b$ and the transformation function specified by the argument $f$. For more details and examples see the chapter on numerical methods in the *Gretl User's Guide*.

See also BFGSmax.

**fft**

Output:      matrix
Argument:    $X$ (matrix)

Discrete real Fourier transform. If the input matrix $X$ has $n$ columns, the output has $2n$ columns, where the real parts are stored in the odd columns and the complex parts in the even ones.

Should it be necessary to compute the Fourier transform on several vectors with the same number of elements, it is numerically more efficient to group them into a matrix rather than invoking fft for each vector separately. See also ffti.

**ffti**

Output:      matrix
Argument:    $X$ (matrix)

Inverse discrete real Fourier transform. It is assumed that $X$ contains $n$ complex column vectors, with the real part in the odd columns and the imaginary part in the even ones, so the total number of columns should be $2n$. A matrix with $n$ columns is returned.

Should it be necessary to compute the inverse Fourier transform on several vectors with the same number of elements, it is numerically more efficient to group them into a matrix rather than invoking ffti for each vector separately. See also fft.

**filter**

| Output: | series |
|---|---|
| Arguments: | $x$ (series) |
| | $a$ (scalar or vector, optional) |
| | $b$ (scalar or vector, optional) |
| | $y0$ (scalar, optional) |

Computes an ARMA-like filtering of the series $x$. The transformation can be written as

$$y_t = \sum_{i=0}^{q} a_i x_{t-i} + \sum_{i=i}^{p} b_i y_{t-i}$$

The two arguments $a$ and $b$ are optional. They may be scalars, vectors or the keyword null.

If $a$ is a scalar, this is used as $a_0$ and implies $q = 0$; if it is a vector of $q + 1$ elements, they contain the coefficients from $a_0$ to $a_q$. If $a$ is null or omitted, this is equivalent to setting $a_0 = 1$ and $q = 0$.

If $b$ is a scalar, this is used as $b_1$ and implies $p = 1$; if it is a vector of $p$ elements, they contain the coefficients from $b_1$ to $b_p$. If $b$ is null or omitted, this is equivalent to setting $B(L) = 1$.

The optional scalar argument $y0$ is taken to represent all values of $y$ prior to the beginning of sample (used only when $p > 0$). If omitted, it is understood to be 0. Pre-sample values of $x$ are always assumed zero.

See also bkfilt, fracdiff, hpfilt, movavg.

Example:

```
nulldata 5
y = filter(index, 0.5, -0.9, 1)
print index y --byobs
```

produces

```
              index          y

1             1        -0.40000
2             2         1.36000
3             3         0.27600
4             4         1.75160
5             5         0.92356
```

**firstobs**

| Output: | scalar |
|---|---|
| Argument: | $y$ (series) |

First non-missing observation for the variable $y$. Note that if some form of subsampling is in effect, the value returned may be smaller than the dollar variable $t1$. See also lastobs.

## floor

Output:        same type as input
Argument:     $y$ (scalar, series or matrix)

Floor function: returns the greatest integer less than or equal to $x$. Note: int and floor differ in their effect for negative arguments: int(-3.5) gives $-3$, while floor(-3.5) gives $-4$.

## fracdiff

Output:        series
Arguments:     $y$ (series)
               $d$ (scalar)

$$\Delta^d y_t = y_t - \sum_{i=1}^{\infty} \psi_i y_{t-i}$$

where

$$\psi_i = \frac{\Gamma(i-d)}{\Gamma(-d)\Gamma(i+1)}$$

Note that in theory fractional differentiation is an infinitely long filter. In practice, presample values of $y_t$ are assumed to be zero.

## gammafun

Output:        same type as input
Argument:     $x$ (scalar, series or matrix)

Returns the gamma function of $x$.

## getenv

Output:        string
Argument:     $s$ (string)

If an environment variable by the name of $s$ is defined, returns the string value of that variable, otherwise returns an empty string. See also ngetenv.

## gini

Output:        scalar
Argument:     $y$ (series)

Returns Gini's inequality index for the series $y$.

## ginv

Output:        matrix
Argument:     $A$ (matrix)

Returns $A^+$, the Moore–Penrose or generalized inverse of $A$, computed via the singular value decomposition.

This matrix has the properties

$$\begin{aligned} AA^+A &= A \\ A^+AA^+ &= A^+ \end{aligned}$$

Moreover, the products $A^+A$ and $AA^+$ are symmetric by construction.

See also inv, svd.

**hdprod**

Output:       matrix
Arguments:    $X$ (matrix)
              $Y$ (matrix)

Horizontal direct product. The two arguments must have the same number of rows, $r$. The return value is a matrix with $r$ rows, in which the $i$-th row is the Kronecker product of the corresponding rows of $X$ and $Y$.

In other words, if $X$ is an $r \times k$ matrix, $Y$ is an $r \times m$ matrix and $Z$ is the result matrix of the horizontal direct product of $X$ times $Y$, then $Z$ will have $r$ rows and $k \cdot m$ columns; moreover,

$$Z_{in} = X_{ij} Y_{il}$$

where $n = (j - 1)m + l$.

As far as we know, there isn't an established name for this operation in matrix algebra. "Horizontal direct product" is the way this operation is called in the GAUSS programming language.

Example: the code

```
A = {1,2,3; 4,5,6}
B = {0,1; -1,1}
C = hdprod(A, B)
```

produces the following matrix:

```
 0    1    0    2    0    3
-4    4   -5    5   -6    6
```

**hpfilt**

Output:       series
Arguments:    $y$ (series)
              *lambda* (scalar, optional)

Returns the cycle component from application of the Hodrick–Prescott filter to series $y$. If the smoothing parameter, *lambda*, is not supplied then a data-based default is used, namely 100 times the square of the periodicity (100 for annual data, 1600 for quarterly data, and so on). See also bkfilt.

**I**

Output:       square matrix
Argument:     $n$ (scalar)

Returns an identity matrix with $n$ rows and columns.

**imaxc**

Output:       row vector
Argument:     $X$ (matrix)

Returns the row indices of the maxima of the columns of $X$.

See also imaxr, iminc, maxc.

**imaxr**

  Output:      column vector
  Argument:   $X$ (matrix)

Returns the column indices of the maxima of the rows of $X$.

See also imaxc, iminr, maxr.

**imhof**

  Output:      scalar
  Arguments:  $M$ (matrix)
             $x$ (scalar)

Computes $\text{Prob}(u'Au < x)$ for a quadratic form in standard normal variates, $u$, using the procedure developed by Imhof (1961).

If the first argument, $M$, is a square matrix it is taken to specify $A$, otherwise if it's a column vector it is taken to be the precomputed eigenvalues of $A$, otherwise an error is flagged.

See also pvalue.

**iminc**

  Output:      row vector
  Argument:   $X$ (matrix)

Returns the row indices of the minima of the columns of $X$.

See also iminr, imaxc, minc.

**iminr**

  Output:      column vector
  Argument:   $X$ (matrix)

Returns the column indices of the mimima of the rows of $X$.

See also iminc, imaxr, minr.

**inbundle**

  Output:      scalar
  Arguments:  $b$ (bundle)
             *key* (string)

Returns 1 if bundle $b$ contains a data-item with name *key*, otherwise 0.

**infnorm**

  Output:      scalar
  Argument:   $X$ (matrix)

Returns the $\infty$-norm of the $r \times c$ matrix $X$, namely,

$$\|X\|_\infty = \max_i \sum_{j=1}^{c} |X_{ij}|$$

See also onenorm.

**inlist**

> Output:        scalar
> Arguments:   *L* (list)
>                      *y* (series)

Returns the (1-based) position of *y* in list *L*, or 0 if *y* is not present in *L*. The second argument may be given as the name of a series or alternatively as an integer ID number.

**int**

> Output:        same type as input
> Argument:    *x* (scalar, series or matrix)

Returns the integer part of *x*, truncating the fractional part. Note: `int` and floor differ in their effect for negative arguments: `int(-3.5)` gives $-3$, while `floor(-3.5)` gives $-4$. See also ceil.

**inv**

> Output:        matrix
> Argument:    *A* (square matrix)

Returns the inverse of *A*. If *A* is singular or not square, an error message is produced and nothing is returned. Note that gretl checks automatically the structure of *A* and uses the most efficient numerical procedure to perform the inversion.

The matrix types gretl checks for are: identity; diagonal; symmetric and positive definite; symmetric but not positive definite; and triangular.

See also ginv, invpd.

**invcdf**

> Output:        same type as input
> Arguments:   *c* (character)
>                      … (see below)
>                      *p* (scalar, series or matrix)

Inverse cumulative distribution function calculator. Returns *x* such that $P(X \leq x) = p$, where the distribution *X* is determined by the character *c*; Between the arguments *c* and *p*, zero or more additional scalar arguments are required to specify the parameters of the distribution, as follows.

| *Distribution* | code, *c* | *Arg* 2 | *Arg* 3 |
|---|---|---|---|
| Standard normal | z, n or N | – | – |
| Gamma | g or G | shape | scale |
| Student's *t* (central) | t | degrees of freedom | – |
| Chi square | c, x or X | degrees of freedom | – |
| Snedecor's *F* | f or F | df (num.) | df (den.) |
| Binomial | b or B | *p* | *n* |
| Poisson | p or P | $\lambda$ | – |

See also cdf, critical, pvalue.

**invmills**

> Output:        same type as input
> Argument:    *x* (scalar, series or matrix)

Returns the inverse Mills ratio at $x$, that is the ratio between the standard normal density and the complement to the standard normal distribution function, both evaluated at $x$.

This function uses a dedicated algorithm which yields greater accuracy compared to calculation using dnorm and cnorm, but the difference between the two methods is appreciable only for very large negative values of $x$.

See also cdf, cnorm, dnorm.

### invpd

| Output: | square matrix |
|---|---|
| Argument: | $A$ (symmetric matrix) |

Returns the inverse of the symmetric, positive definite matrix $A$. This function is slightly faster than inv for large matrices, since no check for symmetry is performed; for that reason it should be used with care.

### irf

| Output: | matrix |
|---|---|
| Arguments: | *target* (scalar) |
| | *shock* (scalar) |
| | *alpha* (scalar between 0 and 1, optional) |

This function is available only when the last model estimated was a VAR or VECM. It returns a matrix containing the estimated response of the *target* variable to an impulse of one standard deviation in the *shock* variable. These variables are identified by their position in the VAR specification: for example, if *target* and *shock* are given as 1 and 3 respectively, the returned matrix gives the response of the first variable in the VAR for a shock to the third variable.

If the optional *alpha* argument is given, the returned matrix has three columns: the point estimate of the responses, followed by the lower and upper limits of a $1 - \alpha$ confidence interval obtained via bootstrapping. (So *alpha* = 0.1 corresponds to 90 percent confidence.) If *alpha* is omitted or set to zero, only the point estimate is provided.

The number of periods (rows) over which the response is traced is determined automatically based on the frequency of the data, but this can be overridden via the set command, as in `set horizon 10`.

### irr

| Output: | scalar |
|---|---|
| Argument: | $x$ (series or vector) |

Returns the Internal Rate of Return for $x$, considered as a sequence of payments (negative) and receipts (positive). See also npv.

### isconst

| Output: | scalar |
|---|---|
| Arguments: | $y$ (series or vector) |
| | *panel-code* (scalar, optional) |

Without the optional second argument, returns 1 if $y$ has a constant value over the current sample range (or over its entire length if $y$ is a vector), otherwise 0.

The second argument is accepted only if the current dataset is a panel and $y$ is a series. In that case a *panel-code* value of 0 calls for a check for time-invariance, while a value of 1 means check for cross-sectional invariance (that is, in each time period the value of $y$ is the same for all groups).

If $y$ is a series, missing values are ignored in checking for constancy.

**islist**

  Output:       scalar
  Argument:   *s* (string)

Returns 1 if *s* is the identifier for a currently defined list, otherwise 0. See also isnull, isseries, isstring.

**isnull**

  Output:       scalar
  Argument:   *s* (string)

Returns 0 if *s* is the identifier for a currently defined object, be it a scalar, a series, a matrix, list or string; otherwise returns 1. See also islist, isseries, isstring.

**kdensity**

  Output:       matrix
  Arguments:  *x* (series)
             *scale* (scalar, optional)
             *control* (scalar, optional)

Computes a kernel density estimate for the series *x*. The returned matrix has two columns, the first holding a set of evenly spaced abscissae and the second the estimated density at each of these points.

The optional *scale* parameter can be used to adjust the degree of smoothing relative to the default of 1.0 (higher values produce a smoother result). The *control* parameter acts as a boolean: 0 (the default) means that the Gaussian kernel is used; a non-zero value switches to the Epanechnikov kernel.

A plot of the results may be obtained using the gnuplot command, as in

```
matrix d = kdensity(x)
gnuplot 2 1 --matrix=d --with-lines
```

**kfilter**

  Output:       scalar
  Arguments:  *&E* (reference to matrix, or `null`)
             *&V* (reference to matrix, or `null`)
             *&S* (reference to matrix, or `null`)
             *&P* (reference to matrix, or `null`)
             *&G* (reference to matrix, or `null`)

Requires that a Kalman filter be set up. Performs a forward, filtering pass and returns 0 on successful completion or 1 if numerical problems are encountered.

The optional matrix arguments can be used to retrieve the following information: *E* gets the matrix of one-step ahead prediction errors and *V* gets the variance matrix for these errors; *S* gets the matrix of estimated values of the state vector and *P* the variance matrix of these estimates; *G* gets the Kalman gain. All of these matrices have *T* rows, corresponding to *T* observations. For the column dimensions and further details see the *Gretl User's Guide*.

See also kalman, ksmooth, ksimul.

**ksimul**

   Output:        matrix
   Arguments:   *v* (matrix)
                   *w* (matrix)
                   *&S* (reference to matrix, or `null`)

Requires that a Kalman filter be set up.  Performs a simulation and returns a matrix holding simulated values of the observable variables.

The argument *v* supplies artificial disturbances for the state transition equation and *w* supplies disturbances for the observation equation, if applicable. The optional argument *S* may be used to retrieve the simulated state vector. For details see the *Gretl User's Guide.*

See also kalman, kfilter, ksmooth.

**ksmooth**

   Output:        matrix
   Argument:   *&P* (reference to matrix, or `null`)

Requires that a Kalman filter be set up.  Performs a backward, smoothing pass and returns a matrix holding smoothed estimates of the state vector. The optional argument *P* may be used to retrieve the MSE of the smoothed state. For details see the *Gretl User's Guide.*

See also kalman, kfilter, ksimul.

**kurtosis**

   Output:        scalar
   Argument:   *x* (series)

Returns the excess kurtosis of the series *x*, skipping any missing observations.

**isseries**

   Output:        scalar
   Argument:   *s* (string)

Returns 1 if *s* is the identifier for a currently defined series, otherwise 0. See also islist, isnull, isstring.

**isstring**

   Output:        scalar
   Argument:   *s* (string)

Returns 1 if *s* is the identifier for a currently defined string, otherwise 0. See also islist, isnull, isseries.

**lags**

   Output:        list
   Arguments:   *p* (scalar)
                   *y* (series or list)

Generates lags 1 to *p* of the series *y*, or if *y* is a list, of all variables in the list. If $p = 0$, the maximum lag defaults to the periodicity of the data; otherwise *p* must be positive.

The generated variables are automatically named according to the template *varname_i* where *varname* is the name of the original series and *i* is the specific lag. The original portion of the name is truncated if necessary, and may be adjusted in case of non-uniqueness in the set of names thus constructed.

**lastobs**

  Output:      scalar
  Argument:   *y* (series)

Last non-missing observation for the variable *y*. Note that if some form of subsampling is in effect, the value returned may be larger than the dollar variable $t2. See also firstobs.

**ldet**

  Output:      scalar
  Argument:   *A* (square matrix)

Returns the natural log of the determinant of *A*, computed via the LU factorization. See also det, rcond.

**ldiff**

  Output:      same type as input
  Argument:   *y* (series or list)

Computes log differences; starting values are set to NA.

When a list is returned, the individual variables are automatically named according to the template ld_*varname* where *varname* is the name of the original series. The name is truncated if necessary, and may be adjusted in case of non-uniqueness in the set of names thus constructed.

See also diff, sdiff.

**lincomb**

  Output:      series
  Arguments:  *L* (list)
             *b* (vector)

Computes a new series as a linear combination of the series in the list *L*. The coefficients are given by the vector *b*, which must have length equal to the number of series in *L*.

See also wmean.

**ljungbox**

  Output:      scalar
  Arguments:  *y* (series)
             *p* (scalar)

Computes the Ljung–Box Q' statistic for the series *y* using lag order *p*. The currently defined sample range is used. The lag order must be greater than or equal to 1 and less than the number of available observations.

This statistic may be referred to the chi-square distribution with *p* degrees of freedom as a test of the null hypothesis that the series *y* is serially independent. See also pvalue.

**lngamma**

  Output:      same type as input
  Argument:   *x* (scalar, series or matrix)

Returns the log of the gamma function of *x*.

**log**

| Output: | same type as input |
|---|---|
| Argument: | $x$ (scalar, series, matrix or list) |

Returns the natural logarithm of $x$; produces NA for non-positive values. Note: `ln` is an acceptable alias for `log`.

When a list is returned, the individual variables are automatically named according to the template `l_varname` where *varname* is the name of the original series. The name is truncated if necessary, and may be adjusted in case of non-uniqueness in the set of names thus constructed.

**log10**

| Output: | same type as input |
|---|---|
| Argument: | $x$ (scalar, series or matrix) |

Returns the base-10 logarithm of $x$; produces NA for non-positive values.

**log2**

| Output: | same type as input |
|---|---|
| Argument: | $x$ (scalar, series or matrix) |

Returns the base-2 logarithm of $x$; produces NA for non-positive values.

**loess**

| Output: | series |
|---|---|
| Arguments: | $y$ (series) |
| | $x$ (series) |
| | $d$ (scalar, optional) |
| | $q$ (scalar, optional) |
| | *robust* (scalar, optional) |

Performs locally-weighted polynomial regression and returns a series holding predicted values of $y$ for each non-missing value of $x$. The method is as described by Cleveland (1979).

The optional arguments $d$ and $q$ specify the order of the polynomial in $x$ and the proportion of the data points to be used in local estimation, respectively. The default values are $d = 1$ and $q = 0.5$. The other acceptable values for $d$ are 0 and 2. Setting $d = 0$ reduces the local regression to a form of moving average. The value of $q$ must be greater than 0 and cannot exceed 1; larger values produce a smoother outcome.

If a non-zero value is given for the *robust* argument the local regressions are iterated twice, with the weights being modified based on the residuals from the previous iteration so as to give less influence to outliers.

See also nadarwat, and in addition see the *Gretl User's Guide* for details on nonparametric methods.

**logistic**

| Output: | same type as input |
|---|---|
| Argument: | $x$ (scalar, series or matrix) |

Returns the logistic function of the argument $x$, that is, $\Lambda(x) = e^x/(1 + e^x)$. If $x$ is a matrix, the function is applied element by element.

**lower**

   Output:      square matrix

   Argument:   $A$ (matrix)

Returns an $n \times n$ lower triangular matrix $B$ for which $B_{ij} = A_{ij}$ if $i \geq j$, and 0 otherwise.

See also upper.

**lrvar**

   Output:      scalar

   Arguments:   $y$ (series or vector)

                   $k$ (scalar)

Returns the long-run variance of $y$, calculated using a Bartlett kernel with window size $k$. If $k$ is negative, `int(T^(1/3))` is used.

In formulae:

$$\hat{\omega}^2(k) = \frac{1}{T} \sum_{t=k}^{T-k} \left[ \sum_{i=-k}^{k} w_i (y_t - \bar{X})(y_{t-i} - \bar{Y}) \right]$$

with

$$w_i = 1 - \frac{|i|}{k+1}$$

**max**

   Output:      scalar or series

   Argument:   $y$ (series or list)

If the argument $y$ is a series, returns the (scalar) maximum of the non-missing observations in the series. If the argument is a list, returns a series each of whose elements is the maximum of the values of the listed variables at the given observation.

See also min, xmax, xmin.

**maxc**

   Output:      row vector

   Argument:   $X$ (matrix)

Returns a row vector containing the maxima of the columns of $X$.

See also imaxc, maxr, minc.

**maxr**

   Output:      column vector

   Argument:   $X$ (matrix)

Returns a column vector containing the maxima of the rows of $X$.

See also imaxc, maxc, minr.

**mcorr**

   Output:      matrix

   Argument:   $X$ (matrix)

Computes a correlation matrix treating each column of $X$ as a variable. See also corr, cov, mcov.

**mcov**

  Output:      matrix
  Argument:   *X* (matrix)

Computes a covariance matrix treating each column of *X* as a variable. See also corr, cov, mcorr.

**mcovg**

  Output:      matrix
  Arguments:   *X* (matrix)
               *u* (vector, optional)
               *w* (vector, optional)
               *p* (scalar)

Returns the matrix covariogram for a $T \times k$ matrix *X* (typically containing regressors), an (optional) *T*-vector *u* (typically containing residuals), an (optional) $(p+1)$-vector of weights *w*, and a scalar lag order *p*, which must be greater than or equal to 0.

The returned matrix is given by

$$\sum_{j=-p}^{p} \sum_{j} w_{|j|} (X_t' u_t u_{t-j} X_{t-j})$$

If *u* is given as `null` the *u* terms are omitted, and if *w* is given as `null` all the weights are taken to be 1.0.

**mean**

  Output:      scalar or series
  Argument:   *x* (series or list)

If *x* is a series, returns the (scalar) sample mean, skipping any missing observations.

If *x* is a list, returns a series *y* such that $y_t$ is the mean of the values of the variables in the list at observation *t*, or `NA` if there are any missing values at *t*.

**meanc**

  Output:      row vector
  Argument:   *X* (matrix)

Returns the means of the columns of *X*. See also meanr, sumc, sdc.

**meanr**

  Output:      column vector
  Argument:   *X* (matrix)

Returns the means of the rows of *X*. See also meanc, sumr.

**median**

  Output:      scalar
  Argument:   *y* (series)

The median of the non-missing observations in series *y*. See also quantile.

**mexp**

  Output:      square matrix

  Argument:   $A$ (square matrix)

Matrix exponential,

$$e^A = \sum_{k=0}^{\infty} \frac{A^k}{k!} = \frac{I}{0!} + \frac{A}{1!} + \frac{A^2}{2!} + \frac{A^3}{3!} + \cdots$$

(This series is sure to converge.) The algorithm used is 11.3.1 from Golub and Van Loan (1996).

**min**

  Output:      scalar or series

  Argument:   $y$ (series or list)

If the argument $y$ is a series, returns the (scalar) minimum of the non-missing observations in the series. If the argument is a list, returns a series each of whose elements is the minimum of the values of the listed variables at the given observation.

See also max, xmax, xmin.

**minc**

  Output:      row vector

  Argument:   $X$ (matrix)

Returns the minima of the columns of $X$.

See also iminc, maxc, minr.

**minr**

  Output:      column vector

  Argument:   $X$ (matrix)

Returns the minima of the rows of $X$.

See also iminr, maxr, minc.

**missing**

  Output:      same type as input

  Argument:   $x$ (scalar, series or list)

Returns a binary variable holding 1 if $x$ is NA. If $x$ is a series, the comparison is done element by element; if $x$ is a list of series, the output is a series with 1 at observations for which at least one series in the list has a missing value, and 0 otherwise.

See also misszero, ok, zeromiss.

**misszero**

  Output:      same type as input

  Argument:   $x$ (scalar or series)

Converts NAs to zeros. If $x$ is a series, the conversion is done element by element. See also missing, ok, zeromiss.

**mlag**

| | |
|---|---|
| Output: | matrix |
| Arguments: | $X$ (matrix) |
| | $p$ (scalar or vector) |
| | $m$ (scalar, optional) |

Shifts up or down the rows of $X$. If $p$ is a positive scalar, the returned matrix $Y$ has typical element $Y_{i,j} = X_{i-p,j}$ for $i \geq p$ and zero otherwise. In other words, the columns of $X$ are shifted down by $p$ rows and the first $p$ rows are filled with the value $m$. If $p$ is a negative number, $X$ is shifted up and the last rows are filled with the value $m$. If $m$ is omitted, it is understood to be zero.

If $p$ is a vector, the above operation is carried out for each element in $p$, joining the resulting matrices horizontally.

**mnormal**

| | |
|---|---|
| Output: | matrix |
| Arguments: | $r$ (scalar) |
| | $c$ (scalar) |

Returns a matrix with $r$ rows and $c$ columns, filled with standard normal pseudo-random variates. See also normal, muniform.

**mols**

| | |
|---|---|
| Output: | matrix |
| Arguments: | $Y$ (matrix) |
| | $X$ (matrix) |
| | &$U$ (reference to matrix, or null) |
| | &$V$ (reference to matrix, or null) |

Returns a $k \times n$ matrix of parameter estimates obtained by OLS regression of the $T \times n$ matrix $Y$ on the $T \times k$ matrix $X$.

If the third argument is not null, the $T \times n$ matrix $U$ will contain the residuals. If the final argument is given and is not null then the $k \times k$ matrix $V$ will contain (a) the covariance matrix of the parameter estimates, if $Y$ has just one column, or (b) $X'X^{-1}$ if $Y$ has multiple columns.

By default, estimates are obtained via Cholesky decomposition, with a fallback to QR decomposition if the columns of $X$ are highly collinear. The use of SVD can be forced via the command `set svd on`.

See also mpols, mrls.

**monthlen**

| | |
|---|---|
| Output: | scalar |
| Arguments: | *month* (scalar) |
| | *year* (scalar) |
| | *weeklen* (scalar) |

Returns the number of (relevant) days in the specified month in the specified year; *weeklen*, which must equal 5, 6 or 7, gives the number of days in the week that should be counted (a value of 6 omits Sundays, and a value of 5 omits both Saturdays and Sundays).

### movavg

| Output: | series |
|---|---|
| Arguments: | *x* (series) |
| | *p* (scalar) |
| | *control* (scalar, optional) |

Depending on the value of the parameter *p*, returns either a simple or an exponentially weighted moving average of the input series *x*.

If $p > 1$, a simple *p*-term moving average is computed, that is, $\frac{1}{p}\sum_{i=0}^{p-1} x_{t-i}$. If a non-zero value is supplied for the optional *control* parameter the MA is centered, otherwise it is "trailing".

If $0 < p < 1$, an exponential moving average is computed: $y_t = px_t + (1-p)y_{t-1}$. By default the output series $y$ is initialized using the first valid value of *x*, but the *control* parameter may be used to specify the number of initial observations that should be averaged to produce $y_0$. A zero value for *control* indicates that all the observations should be used.

### mpols

| Output: | matrix |
|---|---|
| Arguments: | *Y* (matrix) |
| | *X* (matrix) |
| | *&U* (reference to matrix, or `null`) |

Works exactly as mols, except that the calculations are done in multiple precision using the GMP library (assuming this is available).

By default GMP uses 256 bits for each floating point number, but you can adjust this using the environment variable GRETL_MP_BITS, e.g. GRETL_MP_BITS=1024.

### mrandgen

| Output: | matrix |
|---|---|
| Arguments: | *c* (character) |
| | *a* (scalar) |
| | *b* (scalar) |
| Examples: | `matrix mx = mrandgen(u, 0, 100, 50, 1)` |
| | `matrix mt14 = mrandgen(t, 14, 20, 20)` |

Works like randgen except that the return value is a matrix rather than a series. The initial arguments to this function are as described for `randgen`, but they must be followed by two integers to specify the number of rows and columns of the desired random matrix.

The first example above calls for a uniform random column vector of length 50, while the second example specifies a $20 \times 20$ random matrix with drawings from the the *t* distribution with 14 degrees of freedom.

See also mnormal, muniform.

### mread

| Output: | matrix |
|---|---|
| Argument: | *s* (string) |

Reads a matrix from a text file. The string *s* must contain the name of the (plain text) file from which the matrix is to be read. The file in question must conform to the following rules:

- The columns must be separated by spaces or tab characters.

- The decimal separator must be the dot character, ".".

- The first line in the file must contain two integers, separated by a space or a tab, indicating the number of rows and columns, respectively.

Should an error occur (such as the file being badly formatted or inaccessible), an empty matrix is returned.

See also mwrite.

### mreverse

| | |
|---|---|
| Output: | matrix |
| Argument: | *X* (matrix) |

Returns a matrix containing the rows of *X* in reverse order. If you wish to obtain a matrix in which the columns of *X* appear in reverse order you can do:

```
matrix Y = mreverse(X')'
```

### mrls

| | |
|---|---|
| Output: | matrix |
| Arguments: | *Y* (matrix) |
| | *X* (matrix) |
| | *R* (matrix) |
| | *q* (column vector) |
| | *&U* (reference to matrix, or null) |
| | *&V* (reference to matrix, or null) |

Restricted least squares: returns a $k \times n$ matrix of parameter estimates obtained by least-squares regression of the $T \times n$ matrix $Y$ on the $T \times k$ matrix $X$ subject to the linear restriction $RB = q$, where $B$ denotes the stacked coefficient vector. $R$ must have $k * n$ columns; each row of this matrix represents a linear restriction. The number of rows in $q$ must match the number of rows in $R$.

If the fifth argument is not null, the $T \times n$ matrix $U$ will contain the residuals. If the final argument is given and is not null then the $k \times k$ matrix $V$ will hold the restricted counterpart to the matrix $X'X^{-1}$. The variance matrix of the estimates for equation $i$ can be constructed by multiplying the appropriate sub-matrix of $V$ by an estimate of the error variance for that equation.

### mshape

| | |
|---|---|
| Output: | matrix |
| Arguments: | *X* (matrix) |
| | *r* (scalar) |
| | *c* (scalar) |

Rearranges the elements of *X* into a matrix with *r* rows and *c* columns. Elements are read from *X* and written to the target in column-major order. If *X* contains fewer than $k = rc$ elements, the elements are repeated cyclically; otherwise, if *X* has more elements, only the first *k* are used.

See also cols, rows, unvech, vec, vech.

### msortby

| | |
|---|---|
| Output: | matrix |
| Arguments: | *X* (matrix) |
| | *j* (scalar) |

Returns a matrix in which the rows of $X$ are reordered by increasing value of the elements in column $j$.

### muniform

> Output:      matrix
> Arguments:   $r$ (scalar)
>              $c$ (scalar)

Returns a matrix with $r$ rows and $c$ columns, filled with uniform (0,1) pseudo-random variates. Note: the preferred method for generating a scalar uniform r.v. is recasting the output of `muniform` to a scalar, as in

```
scalar x = muniform(1,1)
```

See also mnormal, uniform.

### mwrite

> Output:      scalar
> Arguments:   $X$ (matrix)
>              $s$ (string)

Writes the matrix $X$ to a plain text file named $s$. The file will contain on the first line two integers, separated by a tab character, with the number of rows and columns; on the next lines, the matrix elements in scientific notation, separated by tabs (one line per row).

If file $s$ already exists, it will be overwritten. The return value is 0 on successful completion; if an error occurs, such as the file being unwritable, the return value will be non-zero.

Matrices stored via the `mwrite` command can be easily read by other programs; see the *Gretl User's Guide* for details.

See also mread.

### mxtab

> Output:      matrix
> Arguments:   $x$ (series or vector)
>              $y$ (series or vector)

Returns a matrix holding the cross tabulation of the values contained in $x$ (by row) and $y$ (by column). The two arguments should be of the same type (both series or both column vectors), and because of the typical usage of this function, are assumed to contain integer values only.

See also values.

### nadarwat

> Output:      series
> Arguments:   $y$ (series)
>              $x$ (series)
>              $h$ (scalar)

Returns the Nadaraya–Watson nonparametric estimator of the conditional mean of $y$ given $x$. It returns a series holding the nonparametric estimate of $E(y_i|x_i)$ for each nonmissing element of the series $x$.

$$m(x_i) = \frac{\sum_{j=1}^{n} y_j \cdot K_h(x_i - x_j)}{\sum_{j=1}^{n} K_h(x_i - x_j)}$$

where the kernel function $K_h(\cdot)$ is given by

$$K_h(x) = \exp\left(-\frac{x^2}{2h}\right)$$

for $|x| < \tau$ and zero otherwise.

The argument $h$, known as the *bandwidth*, is a parameter (a positive real number) given by the user. This is usually a small number: larger values of $h$ make $m(x)$ smoother; a popular choice is $n^{-0.2}$. More details are given in the *Gretl User's Guide*.

The scalar $\tau$ is used to prevent numerical problems when the the kernel function is evaluated too far away from zero and is called the trim parameter.

The trim parameter can be adjusted via the `nadarwat_trim` setting, as a multiple of $h$. The default value is 4.

The user may provide a negative value for the bandwidth: this is interpreted as conventional syntax to obtain the leave-one-out estimator, that is a variant of the estimator that does not use the $i$-th observation for evaluating $m(x_i)$. This makes the Nadaraya–Watson estimator more robust numerically and its usage is normally advised when the estimator is computed for inference purposes. Of course, the bandwidth actually used is the absolute value of $h$.

In formulae, the leave-one-out estimator is

$$m(x_i) = \frac{\sum_{j \neq i} y_j \cdot K_h(x_i - x_j)}{\sum_{j \neq i} K_h(x_i - x_j)}$$

**nelem**

  Output:      scalar
  Argument:   $L$ (list)

Returns the number of members in list $L$.

**ngetenv**

  Output:      scalar
  Argument:   $s$ (string)

If an environment variable by the name of $s$ is defined and has a numerical value, returns that value; otherwise returns NA. See also getenv.

**nobs**

  Output:      scalar
  Argument:   $y$ (series)

Returns the number of non-missing observations for the variable $y$ in the currently selected sample.

**normal**

  Output:      series
  Arguments:   $\mu$ (scalar)
                 $\sigma$ (scalar)

Generates a series of Gaussian pseudo-random variates with mean $\mu$ and standard deviation $\sigma$. If no arguments are supplied, standard normal variates $N(0,1)$ are produced. The values are produced using the Ziggurat method (Marsaglia and Tsang, 2000).

See also randgen, mnormal, muniform.

**npv**

| Output: | scalar |
|---|---|
| Arguments: | *x* (series or vector) |
| | *r* (scalar) |

Returns the Net Present Value of *x*, considered as a sequence of payments (negative) and receipts (positive), evaluated at annual discount rate *r*. The first value is taken as dated "now" and is not discounted. To emulate an NPV function in which the first value is discounted, prepend zero to the input sequence.

Supported data frequencies are annual, quarterly, monthly, and undated (undated data are treated as if annual).

See also irr.

**NRmax**

| Output: | scalar |
|---|---|
| Arguments: | *b* (vector) |
| | *f* (function call) |
| | *g* (function call, optional) |
| | *h* (function call, optional) |

Numerical maximization via the Newton–Raphson method. The vector *b* should hold the initial values of a set of parameters, and the argument *f* should specify a call to a function that calculates the (scalar) criterion to be maximized, given the current parameter values and any other relevant data. If the object is in fact minimization, this function should return the negative of the criterion. On successful completion, `NRmax` returns the maximized value of the criterion, and *b* holds the parameter values which produce the maximum.

The optional third and fourth arguments provide means of supplying analytical derivatives and an analytical (negative) Hessian, respectively. The functions referenced by *g* and *h* must take as their first argument a pre-defined matrix that is of the correct size to contain the gradient or Hessian, respectively, given in pointer form. They also must take the parameter vector as an argument (in pointer form or otherwise). Other arguments are optional. If either or both of the optional arguments are omitted, a numerical approximation is used.

For more details and examples see the chapter on numerical methods in the *Gretl User's Guide*. See also BFGSmax, fdjac.

**nullspace**

| Output: | matrix |
|---|---|
| Argument: | *A* (matrix) |

Computes the right nullspace of *A*, via the singular value decomposition: the result is a matrix *B* such that *AB* = [0], except when *A* has full column rank, in which case an empty matrix is returned. Otherwise, if *A* is $m \times n$, *B* will be an $n \times (n - r)$ matrix, where *r* is the rank of *A*.

See also rank, svd.

**obs**

| Output: | series |
|---|---|

Returns a series of consecutive integers, setting 1 at the start of the dataset. Note that the result is invariant to subsampling. This function is especially useful with time-series datasets. Note: you can write `t` instead of `obs` with the same effect.

See also obsnum.

**obslabel**

  Output:        string
  Argument:     *t* (scalar)

Returns the observation label for observation *t*, where *t* is a 1-based index. The inverse function is provided by obsnum.

**obsnum**

  Output:        scalar
  Argument:     *s* (string)

Returns an integer corresponding to the observation specified by the string *s*. Note that the result is invariant to subsampling. This function is especially useful with time-series datasets. For example, the following code

```
open denmark
k = obsnum(1980:1)
```

yields `k = 25`, indicating that the first quarter of 1980 is the 25th observation in the `denmark` dataset.

See also obs, obslabel.

**ok**

  Output:        see below
  Argument:     *x* (scalar, series, matrix or list)

If *x* is a scalar, returns 1 if *x* is not NA, otherwise 0. If *x* is a series, returns a series with value 1 at observations with non-missing values and zeros elsewhere. If *x* is a list, the output is a series with 0 at observations for which at least one series in the list has a missing value, and 1 otherwise.

If *x* is a matrix the behavior is a little different, since matrices cannot contain NAs: the function returns a matrix of the same dimensions as *x*, with 1s in positions corresponding to finite elements of *x* and 0s in positions where the elements are non-finite (either infinities or not-a-number, as per the IEEE 754 standard).

See also missing, misszero, zeromiss. But note that these functions are not applicable to matrices.

**onenorm**

  Output:        scalar
  Argument:     *X* (matrix)

Returns the 1-norm of the $r \times c$ matrix *X*:

$$\|X\|_1 = \max_j \sum_{i=1}^{r} |X_{ij}|$$

See also infnorm, rcond.

**ones**

  Output:         matrix
  Arguments:    *r* (scalar)
                        *c* (scalar)

Outputs a matrix with $r$ rows and $c$ columns, filled with ones.

See also seq, zeros.

**orthdev**

| | |
|---|---|
| Output: | series |
| Argument: | $y$ (series) |

Only applicable if the currently open dataset has a panel structure. Computes the forward orthogonal deviations for variable $y$, that is

$$\tilde{y}_{i,t} = \sqrt{\frac{T_i - t}{T_i - t + 1}} \left( y_{i,t} - \frac{1}{T_i - t} \sum_{s=t+1}^{T_i} y_{i,s} \right)$$

This transformation is sometimes used instead of differencing to remove individual effects from panel data. For compatibility with first differences, the deviations are stored one step ahead of their true temporal location (that is, the value at observation $t$ is the deviation that, strictly speaking, belongs at $t - 1$). That way one loses the first observation in each time series, not the last.

See also diff.

**pdf**

| | |
|---|---|
| Output: | same type as input |
| Arguments: | $c$ (character) |
| | ... (see below) |
| | $x$ (scalar, series or matrix) |
| Examples: | `f1 = pdf(N, -2.5)` |
| | `f2 = pdf(X, 3, y)` |
| | `f3 = pdf(W, shape, scale, y)` |

Probability density function calculator. Returns the density at $x$ of the distribution identified by the code $c$. See cdf for details of the required (scalar) arguments. The distributions supported by the pdf function are the normal, Student's $t$, chi-square, $F$, Gamma, Weibull, Generalized Error, Binomial and Poisson. Note that for the Binomial and the Poisson what's calculated is in fact the probability mass at the specified point.

For the normal distribution, see also dnorm.

**pergm**

| | |
|---|---|
| Output: | matrix |
| Arguments: | $x$ (series or vector) |
| | *bandwidth* (scalar, optional) |

If only the first argument is given, computes the sample periodogram for the given series or vector. If the second argument is given, computes an estimate of the spectrum of $x$ using a Bartlett lag window of the given bandwidth, up to a maximum of half the number of observations ($T/2$).

Returns a matrix with two columns and $T/2$ rows: the first column holds the frequency, $\omega$, from $2\pi/T$ to $\pi$, and the second the corresponding spectral density.

**pmax**

| | |
|---|---|
| Output: | series |
| Arguments: | $y$ (series) |
| | *mask* (series, optional) |

Only applicable if the currently open dataset has a panel structure. Returns the per-unit maximum for variable $y$.

If the optional second argument is provided then observations for which the value of *mask* is zero are ignored.

See also pmin, pmean, pnobs, psd, pxsum.

### pmean

   Output:      series
  Arguments:   *y* (series)
               *mask* (series, optional)

Only applicable if the currently open dataset has a panel structure. Computes the per-unit mean for variable $y$; that is,

$$\bar{y}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} y_{i,t}$$

where $T_i$ is the number of valid observations for unit $i$.

If the optional second argument is provided then observations for which the value of *mask* is zero are ignored.

See also pmax, pmin, pnobs, psd, pxsum, pshrink.

### pmin

   Output:      series
  Arguments:   *y* (series)
               *mask* (series, optional)

Only applicable if the currently open dataset has a panel structure. Returns the per-unit mimimum for variable $y$.

If the optional second argument is provided then observations for which the value of *mask* is zero are ignored.

See also pmax, pmean, pnobs, psd.

### pnobs

   Output:      series
  Arguments:   *y* (series)
               *mask* (series, optional)

Only applicable if the currently open dataset has a panel structure. Returns for each unit the number of non-missing cases for the variable $y$.

If the optional second argument is provided then observations for which the value of *mask* is zero are ignored.

See also pmax, pmin, pmean, psd.

### polroots

   Output:      matrix
  Argument:   *a* (vector)

Finds the roots of a polynomial. If the polynomial is of degree $p$, the vector $a$ should contain $p + 1$ coefficients in ascending order, i.e. starting with the constant and ending with the coefficient on $x^p$.

If all the roots are real they are returned in a column vector of length $p$, otherwise a $p \times 2$ matrix is returned, the real parts in the first column and the imaginary parts in the second.

### polyfit

| | |
|---|---|
| Output: | series |
| Arguments: | $y$ (series) |
| | $q$ (scalar) |

Fits a polynomial trend of order $q$ to the input series $y$ using the method of orthogonal polynomials. The series returned holds the fitted values.

### princomp

| | |
|---|---|
| Output: | matrix |
| Arguments: | $X$ (matrix) |
| | $p$ (scalar) |
| | *covmat* (scalar, optional) |

Let the matrix $X$ be $T \times k$, containing $T$ observations on $k$ variables. The argument $p$ must be a positive integer less than or equal to $k$. This function returns a $T \times p$ matrix, $P$, holding the first $p$ principal components of $X$.

The optional third argument acts as a boolean switch: if it is non-zero the principal components are computed on the basis of the covariance matrix of the columns of $X$ (the default is to use the correlation matrix).

The elements of $P$ are computed as

$$P_{tj} = \sum_{i=1}^{k} Z_{ti}\, v_i^{(j)}$$

where $Z_{ti}$ is the standardized value of variable $i$ at observation $t$, $Z_{ti} = (X_{ti} - \bar{X}_i)/\hat{\sigma}_i$, and $v^{(j)}$ is the $j$th eigenvector of the correlation (or covariance) matrix of the $X_i$s, with the eigenvectors ordered by decreasing value of the corresponding eigenvalues.

See also eigensym.

### psd

| | |
|---|---|
| Output: | series |
| Arguments: | $y$ (series) |
| | *mask* (series, optional) |

Only applicable if the currently open dataset has a panel structure. Computes the per-unit sample standard deviation for variable $y$, that is

$$\sigma_i = \sqrt{\frac{1}{T_i - 1} \sum_{t=1}^{T_i} (y_{i,t} - \bar{y}_i)^2}$$

The above formula holds for $T_i \geq 2$, where $T_i$ is the number of valid observations for unit $i$; if $T_i = 0$, NA is returned; if $T_i = 1$, 0 is returned.

If the optional second argument is provided then observations for which the value of *mask* is zero are ignored.

Note: this function makes it possible to check whether a given variable (say, X) is time-invariant via the condition `max(psd(X)) = 0`.

See also pmax, pmin, pmean, pnobs.

### psdroot

Output:     square matrix
Argument:   *A* (symmetric matrix)

Performs a generalized variant of the Cholesky decomposition of the matrix *A*, which must be positive semidefinite (but which may be singular). If the input matrix is not square an error is flagged, but symmetry is assumed and not tested; only the lower triangle of *A* is read. The result is a lower-triangular matrix *L* which satisfies $A = LL'$. Indeterminate elements in the solution are set to zero.

For the case where *A* is positive definite, see cholesky.

### pshrink

Output:     matrix
Argument:   *y* (series)

### pvalue

Output:       same type as input
Arguments:    *c* (character)
              ... (see below)
              *x* (scalar, series or matrix)
Examples:     p1 = pvalue(z, 2.2)
              p2 = pvalue(X, 3, 5.67)
              p2 = pvalue(F, 3, 30, 5.67)

*P*-value calculator. Returns $P(X > x)$, where the distribution *X* is determined by the character *c*. Between the arguments *c* and *x*, zero or more additional arguments are required to specify the parameters of the distribution; see cdf for details. The distributions supported by the pval function are the standard normal, *t*, Chi square, *F*, gamma, binomial, Poisson, Weibull and Generalized Error.

See also critical, invcdf, urcpval, imhof.

### pxsum

Output:       series
Arguments:    *y* (series)
              *mask* (series, optional)

Only applicable if the currently open dataset has a panel structure. Computes the cross-sectional sum for variable *y* in each period; that is,

$$\tilde{y}_t = \sum_{i=1}^{N} y_{i,t}$$

where *N* is the number of cross-sectional units.

If the optional second argument is provided then observations for which the value of *mask* is zero are ignored.

Note that this function works in a different dimension from the pmean function.

### qform

Output:       matrix
Arguments:    *x* (matrix)
              *A* (symmetric matrix)

Computes the quadratic form $Y = xAx'$. Using this function instead of ordinary matrix multiplication guarantees more speed and better accuracy. If $x$ and $A$ are not conformable, or $A$ is not symmetric, an error is returned.

### qnorm

Output:       same type as input

Argument:    $x$ (scalar, series or matrix)

Returns quantiles for the standard normal distribution. If $x$ is not between 0 and 1, NA is returned. See also cnorm, dnorm.

### qrdecomp

Output:       matrix

Arguments:   $X$ (matrix)

$\quad\quad\quad\quad$ &R (reference to matrix, or null)

Computes the QR decomposition of an $m \times n$ matrix $X$, that is $X = QR$ where $Q$ is an $m \times n$ orthogonal matrix and $R$ is an $n \times n$ upper triangular matrix. The matrix $Q$ is returned directly, while $R$ can be retrieved via the optional second argument.

See also eigengen, eigensym, svd.

### quantile

Output:       scalar or matrix

Arguments:   $y$ (series or matrix)

$\quad\quad\quad\quad$ $p$ (scalar between 0 and 1)

If $y$ is a series, returns the $p$-quantile for the series. For example, when $p = 0.5$, the median is returned.

If $y$ is a matrix, returns a row vector containing the $p$-quantiles for the columns of $y$; that is, each column is treated as a series.

In addition, for matrix $y$ an alternate form of the second argument is supported: $p$ may be given as a vector. In that case the return value is an $m \times n$ matrix, where $m$ is the number of elements in $p$ and $n$ is the number of columns in $y$.

For a series of length $n$, the $p$-quantile, $q$, is defined as:

$$q = y_{[k]} + (n \cdot p - k)(y_{[k+1]} - y_{[k]})$$

where $k$ is the integer part of $n \cdot p$ and $y_{[i]}$ is the $i$-th element of the series when sorted from smallest to largest.

### randgen

Output:       series

Arguments:   $c$ (character)

$\quad\quad\quad\quad$ $a$ (scalar or series)

$\quad\quad\quad\quad$ $b$ (scalar or series)

Examples:    ```series x = randgen(u, 0, 100)```

$\quad\quad\quad\quad$ ```series t14 = randgen(t, 14)```

$\quad\quad\quad\quad$ ```series y = randgen(B, 0.6, 30)```

$\quad\quad\quad\quad$ ```series g = randgen(G, 1, 1)```

$\quad\quad\quad\quad$ ```series P = randgen(P, mu)```

All-purpose random number generator. The parameter $c$ is a character, which specifies from which distribution the pseudo-random numbers should be drawn. The arguments $a$ and (in

some cases) *b* provide the parameters of the selected distribution. If these are given as scalars the output series is identically distributed; if a series is given for *a* or *b* the distribution is conditional on the parameter value at each observation.

Specifics are given below: the character codes for each distribution are shown in parentheses, followed by the interpretation of the argument *a* and, where applicable, *b*.

| Distribution | *c* | *a* | *b* |
|---|---|---|---|
| Uniform (continuous) | u or U | minimum | maximum |
| Uniform (discrete) | i | minimum | maximum |
| Normal | z, n or N | mean | standard deviation |
| Student's *t* | t | degrees of freedom | – |
| Chi square | c, x or X | degrees of freedom | – |
| Snedecor's *F* | f or F | df (num.) | df (den.) |
| Gamma | g or G | shape | scale |
| Binomial | b or B | $p$ | $n$ |
| Poisson | p or P | mean | – |
| Weibull | w or W | shape | scale |
| Generalized Error | e or E | shape | – |

See also normal, uniform, mrandgen.

## randint

  Output:        scalar
  Arguments:   *min* (scalar)
                  *max* (scalar)

Returns a pseudo-random integer in the closed interval [*min*, *max*]. See also randgen.

## rank

  Output:        scalar
  Argument:    *X* (matrix)

Returns the rank of *X*, numerically computed via the singular value decomposition. See also svd.

## ranking

  Output:        same type as input
  Argument:    *y* (series or vector)

Returns a series or vector with the ranks of *y*. The rank for observation $i$ is the number of elements that are less than $y_i$ plus one half the number of elements that are equal to $y_i$. (Intuitively, you may think of chess points, where victory gives you one point and a draw gives you half a point.) One is added so the lowest rank is 1 instead of 0.

Formally,

$$\text{rank}(y_i) = 1 + \sum_{j \neq i} \left[ I(y_j < y_i) + 0.5 \cdot I(y_j = y_i) \right]$$

where *I* denotes the indicator function.

See also sort, sortby.

**rcond**

Output:        scalar

Argument:    *A* (square matrix)

Returns the reciprocal condition number for *A* with respect to the 1-norm.  In many circumstances, this is a better measure of the sensitivity of *A* to numerical operations such as inversion than the determinant.

Given that *A* is non-singular, we may define

$$\kappa(A) = ||A||_1 \cdot ||A^{-1}||_1$$

This function returns $\kappa(A)^{-1}$.

See also det, ldet, onenorm.

**readfile**

Output:        string

Argument:    *fname* (string)

If a file by the name of *fname* exists and is readable, returns a string containing the content of this file, otherwise flags an error.

Also see the sscanf function.

**replace**

Output:        same type as input

Arguments:   *x* (series or matrix)
               *find* (scalar or vector)
               *subst* (scalar or vector)

Replaces each element of *x* equal to the *i*-th element of *find* with the corresponding element of *subst*.

If *find* is a scalar, *subst* must also be a scalar.  If *find* and *subst* are both vectors, they must have the same number of elements.  But if *find* is a vector and *subst* a scalar, then all matches will be replaced by *subst*.

Example:

```
a = {1,2,3;3,4,5}
find = {1,3,4}
subst = {-1,-8, 0}
b = replace(a, find, subst)
print a b
```

produces

```
a (2 x 3)

  1   2   3
  3   4   5

b (2 x 3)

 -1   2  -8
 -8   0   5
```

**resample**

  Output:       same type as input

  Arguments:   $x$ (series or matrix)

               $b$ (scalar, optional)

Resamples from $x$ with replacement. In the case of a series argument, each value of the returned series, $y_t$, is drawn from among all the values of $x_t$ with equal probability. When a matrix argument is given, each row of the returned matrix is drawn from the rows of $x$ with equal probability.

The optional argument $b$ represents the block length for resampling by moving blocks. If this argument is given it should be a positive integer greater than or equal to 2. The effect is that the output is composed by random selection with replacement from among all the possible contiguous sequences of length $b$ in the input. (In the case of matrix input, this means contiguous rows.) If the length of the data is not an integer multiple of the block length, the last selected block is truncated to fit.

**round**

  Output:       same type as input

  Argument:   $x$ (scalar, series or matrix)

Rounds to the nearest integer. Note that when $x$ lies halfway between two integers, rounding is done "away from zero", so for example 2.5 rounds to 3, but `round(-3.5)` gives $-4$. This is a common convention in spreadsheet programs, but other software may yield different results. See also ceil, floor, int.

**rownames**

  Output:       scalar

  Arguments:   $M$ (matrix)

               $s$ (named list or string)

Attaches names to the rows of the $m \times n$ matrix $M$. If $s$ is a named list, the row names are copied from the names of the variables; the list must have $m$ members. If $s$ is a string, it should contain $m$ space-separated sub-strings. The return value is 0 on successful completion, non-zero on error. See also colnames.

**rows**

  Output:       scalar

  Argument:   $X$ (matrix)

Number of rows of the matrix $X$. See also cols, mshape, unvech, vec, vech.

**sd**

  Output:       scalar or series

  Argument:   $x$ (series or list)

If $x$ is a series, returns the (scalar) sample standard deviation, skipping any missing observations.

If $x$ is a list, returns a series $y$ such that $y_t$ is the sample standard deviation of the values of the variables in the list at observation $t$, or NA if there are any missing values at $t$.

See also var.

**sdc**

| Output: | row vector |
|---|---|
| Arguments: | *X* (matrix) |
| | *df* (scalar, optional) |

Returns the standard deviations of the columns of *X*. If *df* is positive it is used as the divisor for the column variances, otherwise the divisor is the number of rows in *X* (that is, no degrees of freedom correction is applied). See also meanc, sumc.

**sdiff**

| Output: | same type as input |
|---|---|
| Argument: | *y* (series or list) |

Computes seasonal differences: $y_t - y_{t-k}$, where $k$ is the periodicity of the current dataset (see $pd). Starting values are set to NA.

When a list is returned, the individual variables are automatically named according to the template **sd_***varname* where *varname* is the name of the original series. The name is truncated if necessary, and may be adjusted in case of non-uniqueness in the set of names thus constructed.

See also diff, ldiff.

**selifc**

| Output: | matrix |
|---|---|
| Arguments: | *A* (matrix) |
| | *b* (row vector) |

Selects from *A* only the columns for which the corresponding element of *b* is non-zero. *b* must be a row vector with the same number of columns as *A*.

See also selifr.

**selifr**

| Output: | matrix |
|---|---|
| Arguments: | *A* (matrix) |
| | *b* (column vector) |

Selects from *A* only the rows for which the corresponding element of *b* is non-zero. *b* must be a column vector with the same number of rows as *A*.

See also selifc, trimr.

**seq**

| Output: | row vector |
|---|---|
| Arguments: | *a* (scalar) |
| | *b* (scalar) |
| | *k* (scalar, optional) |

Given only two arguments, returns a row vector filled with consecutive integers, with *a* as first element and *b* last. If *a* is greater than *b* the sequence will be decreasing. If either argument is not integral its fractional part is discarded.

If the third argument is given, returns a row vector containing a sequence of integers starting with *a* and incremented (or decremented, if *a* is greater than *b*) by *k* at each step. The final value is the largest member of the sequence that is less than or equal to *b* (or mutatis mutandis for *a* greater than *b*). The argument *k* must be positive; if it is not integral its fractional part is discarded.

See also ones, zeros.

**simann**

  Output:      scalar
  Arguments:  *b* (vector)
               *f* (function call)
               *maxit* (scalar, optional)

Implements simulated annealing, which may be helpful in improving the initialization for a numerical optimization problem.

The first argument holds the intial value of a parameter vector and the second argument specifies a function call which returns the (scalar) value of the maximand. The optional third argument specifies the maximum number of iterations (which defaults to 1024). On successful completion, `simann` returns the final value of the maximand.

For more details and an example see the chapter on numerical methods in the *Gretl User's Guide*. See also BFGSmax, NRmax.

**sin**

  Output:      same type as input
  Argument:   *x* (scalar, series or matrix)

Returns the sine of *x*. See also cos, tan, atan.

**sinh**

  Output:      same type as input
  Argument:   *x* (scalar, series or matrix)

Returns the hyperbolic sine of *x*.

$$\sinh x = \frac{e^x - e^{-x}}{2}$$

See also asinh, cosh, tanh.

**skewness**

  Output:      scalar
  Argument:   *x* (series)

Returns the skewness value for the series *x*, skipping any missing observations.

**sort**

  Output:      same type as input
  Argument:   *x* (series or vector)

Sorts *x* in ascending order, skipping observations with missing values when *x* is a series. See also dsort, values. For matrices specifically, see msortby.

**sortby**

  Output:      series
  Arguments:  *y1* (series)
               *y2* (series)

Returns a series containing the elements of *y2* sorted by increasing value of the first argument, *y1*. See also sort, ranking.

**sqrt**

Output:       same type as input
Argument:    *x* (scalar, series or matrix)

Returns the positive square root of *x*; produces NA for negative values.

Note that if the argument is a matrix the operation is performed element by element and, since matrices cannot contain NA, negative values generate an error. For the "matrix square root" see cholesky.

**sscanf**

Output:       scalar
Arguments:    *src* (string)
              *format* (string)
              . . . (see below)

Reads values from *src* under the control of *format* and assigns these values to one or more trailing arguments, indicated by the dots above. Returns the number of values assigned. This is a simplifed version of the sscanf function in the C programming language.

*src* may be either a literal string, enclosed in double quotes, or the name of a predefined string variable. *format* is defined similarly to the format string in printf (more on this below). *args* should be a comma-separated list containing the names of pre-defined variables: these are the targets of conversion from *src*. (For those used to C: one can prefix the names of numerical variables with & but this is not required.)

Literal text in *format* is matched against *src*. Conversion specifiers start with %, and recognized conversions include %f, %g or %lf for floating-point numbers; %d for integers; %s for strings; and %m for matrices. You may insert a positive integer after the percent sign: this sets the maximum number of characters to read for the given conversion (or the maximum number of rows in the case of matrix conversion). Alternatively, you can insert a literal * after the percent to suppress the conversion (thereby skipping any characters that would otherwise have been converted for the given type). For example, %3d converts the next 3 characters in *source* to an integer, if possible; %*g skips as many characters in *source* as could be converted to a single floating-point number.

Matrix conversion works thus: the scanner reads a line of input and counts the (space- or tab-separated) number of numeric fields. This defines the number of columns in the matrix. By default, reading then proceeds for as many lines (rows) as contain the same number of numeric columns, but the maximum number of rows to read can be limited as described above.

In addition to %s conversion for strings, a simplified version of the C format %N[*chars*] is available. In this format *N* is the maximum number of characters to read and *chars* is a set of acceptable characters, enclosed in square brackets: reading stops if *N* is reached or if a character not in *chars* is encountered. The function of *chars* can be reversed by giving a circumflex, ∧, as the first character; in that case reading stops if a character in the given set is found. (Unlike C, the hyphen does not play a special role in the *chars* set.)

If the source string does not (fully) match the format, the number of conversions may fall short of the number of arguments given. This is not in itself an error so far as gretl is concerned. However, you may wish to check the number of conversions performed; this is given by the return value.

Some examples follow:

```
scalar x
scalar y
sscanf("123456", "%3d%3d", x, y)

sprintf S, "1 2 3 4\n5 6 7 8"
S
matrix m
```

```
sscanf(S, "%m", m)
print m
```

**sst**

| Output: | scalar |
|---|---|
| Argument: | *y* (series) |

Returns the sum of squared deviations from the mean for the non-missing observations in series *y*. See also var.

**strlen**

| Output: | scalar |
|---|---|
| Argument: | *s* (string) |

Returns the number of characters in *s*.

**strncmp**

| Output: | scalar |
|---|---|
| Arguments: | *s1* (string) |
| | *s2* (string) |
| | *n* (scalar, optional) |

Compares the two string arguments and returns an integer less than, equal to, or greater than zero if *s1* is found, respectively, to be less than, to match, or be greater than *s2*, up to the first *n* characters. If *n* is omitted the comparison proceeds as far as possible.

Note that if you just want to compare two strings for equality, that can be done without using a function, as in if (s1 == s2) ...

**strsplit**

| Output: | string |
|---|---|
| Arguments: | *s* (string) |
| | *i* (scalar) |

Returns space-separated element *i* from the string *s*. The index *i* is 1-based, and it is an error if *i* is less than 1. In case *s* contains no spaces and *i* equals 1, a copy of the entire input string is returned; otherwise, in case *i* exceeds the number of space-separated elements an empty string is returned.

**strstr**

| Output: | string |
|---|---|
| Arguments: | *s1* (string) |
| | *s2* (string) |

Searches *s1* for an occurrence of the string *s2*. If a match is found, returns a copy of the portion of *s1* that starts with *s2*, otherwise returns an empty string.

**strsub**

| Output: | string |
|---|---|
| Arguments: | *s* (string) |
| | *find* (string) |
| | *subst* (string) |

Returns a copy of *s* in which all occurrences of *find* are replaced by *subst*.

### sum

  Output:       scalar or series
  Argument:   *x* (series or list)

If *x* is a series, returns the (scalar) sum of the non-missing observations in *x*. See also sumall.

If *x* is a list, returns a series $y$ such that $y_t$ is the sum of the values of the variables in the list at observation $t$, or NA if there are any missing values at $t$.

### sumall

  Output:       scalar
  Argument:   *x* (series)

Returns the sum of the observations of *x* over the current sample range, or NA if there are any missing values.

### sumc

  Output:       row vector
  Argument:   *X* (matrix)

Returns the sums of the columns of *X*. See also meanc, sumr.

### sumr

  Output:       column vector
  Argument:   *X* (matrix)

Returns the sums of the rows of *X*. See also meanr, sumc.

### svd

  Output:       row vector
  Arguments:  *X* (matrix)
              *&U* (reference to matrix, or null)
              *&V* (reference to matrix, or null)

Performs the singular values decomposition of the $r \times c$ matrix $X$:

$$X = U \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_{n,} \end{bmatrix} V$$

where $n = \min(r, c)$. $U$ is $r \times n$ and $V$ is $n \times c$, with $U'U = I$ and $VV' = I$.

The singular values are returned in a row vector. The left and/or right singular vectors $U$ and $V$ may be obtained by supplying non-null values for arguments 2 and 3, respectively. For any matrix A, the code

```
s = svd(A, &U, &V)
B = (U .* s) * V
```

should yield B identical to A (apart from machine precision).

See also eigengen, eigensym, qrdecomp.

**tan**

  Output:      same type as input
  Argument:    $x$ (scalar, series or matrix)

Returns the tangent of $x$.

**tanh**

  Output:      same type as input
  Argument:    $x$ (scalar, series or matrix)

Returns the hyperbolic tangent of $x$.

$$\tanh x = \frac{e^{2x} - 1}{e^{2x} + 1}$$

See also atanh, cosh, sinh.

**toepsolv**

  Output:      column vector
  Arguments:   $c$ (vector)
               $r$ (vector)
               $b$ (vector)

Solves a Toeplitz system of linear equations, that is $Tx = b$ where $T$ is a square matrix whose element $T_{i,j}$ equals $c_{i-j}$ for $i \geq j$ and $r_{j-i}$ for $i \leq j$. Note that the first elements of $c$ and $r$ must be equal, otherwise an error is returned. Upon successful completion, the function returns the vector $x$.

The algorithm used here takes advantage of the special structure of the matrix $T$, which makes it much more efficient than other unspecialized algorithms, especially for large problems. Warning: in certain cases, the function may spuriously issue a singularity error when in fact the matrix $T$ is nonsingular; this problem, however, cannot arise when $T$ is positive definite.

**tolower**

  Output:      string
  Argument:    $s$ (string)

Returns a copy of $s$ in which any upper-case characters are converted to lower case.

**tr**

  Output:      scalar
  Argument:    $A$ (square matrix)

Returns the trace of the square matrix $A$, that is, the sum of its diagonal elements. See also diag.

**transp**

  Output:      matrix
  Argument:    $X$ (matrix)

Returns the transpose of $X$. Note: this is rarely used; in order to get the transpose of a matrix, in most cases you can just use the prime operator: X'.

**trimr**

  Output:       matrix
  Arguments:   *X* (matrix)
              *ttop* (scalar)
              *tbot* (scalar)

Returns a matrix that is a copy of *X* with *ttop* rows trimmed at the top and *tbot* rows trimmed at the bottom. The latter two arguments must be non-negative, and must sum to less than the total rows of *X*.

See also selifr.

**uniform**

  Output:       series
  Arguments:   *a* (scalar)
              *b* (scalar)

Generates a series of uniform pseudo-random variates in the interval $(a, b)$, or, if no arguments are supplied, in the interval $(0,1)$. The algorithm used by default is the SIMD-oriented Fast Mersenne Twister developed by Saito and Matsumoto (2008).

See also randgen, normal, mnormal, muniform.

**uniq**

  Output:       column vector
  Argument:   *x* (series or vector)

Returns a vector containing the distinct elements of *x*, not sorted but in their order of appearance. See values for a variant that sorts the elements.

**unvech**

  Output:       square matrix
  Argument:   *v* (vector)

Returns an $n \times n$ symmetric matrix obtained by rearranging the elements of $v$. The number of elements in $v$ must be a triangular integer — i.e., a number $k$ such that an integer $n$ exists with the property $k = n(n + 1)/2$. This is the inverse of the function vech.

See also mshape, vech.

**upper**

  Output:       square matrix
  Argument:   *A* (square matrix)

Returns an $n \times n$ upper triangular matrix $B$ for which $B_{ij} = A_{ij}$ if $i \leq j$ and 0 otherwise.

See also lower.

**urcpval**

  Output:       scalar
  Arguments:   *tau* (scalar)
              *n* (scalar)
              *niv* (scalar)
              *itv* (scalar)

*P*-values for the test statistic from the Dickey–Fuller unit-root test and the Engle–Granger cointegration test, as per MacKinnon (1996).

The arguments are as follows: *tau* denotes the test statistic; *n* is the number of observations (or 0 for an asymptotic result); *niv* is the number of potentially cointegrated variables when testing for cointegration (or 1 for a univariate unit-root test); and *itv* is a code for the model specification: 1 for no constant, 2 for constant included, 3 for constant and linear trend, 4 for constant and quadratic trend.

Note that if the test regression is "augmented" with lags of the dependent variable, then you should give an *n* value of 0 to get an asymptotic result.

See also pvalue.

### values

  Output:       column vector
  Argument:     *x* (series or vector)

Returns a vector containing the distinct elements of *x* sorted in ascending order. If you wish to truncate the values to integers before applying this function, use the expression `values(int(x))`.

See also uniq, dsort, sort.

### var

  Output:       scalar or series
  Argument:     *x* (series or list)

If *x* is a series, returns the (scalar) sample variance, skipping any missing observations.

If *x* is a list, returns a series $y$ such that $y_t$ is the sample variance of the values of the variables in the list at observation $t$, or NA if there are any missing values at $t$.

In each case the sum of squared deviations from the mean is divided by $(n - 1)$ for $n > 1$. Otherwise the variance is given as zero if $n = 1$, or as NA if $n = 0$.

See also sd.

### varname

  Output:       string
  Argument:     *v* (scalar or list)

If given a scalar argument, returns the name of the variable with ID number *v*, or generates an error if there is no such variable.

If given a list argument, returns a string containing the names of the variables in the list, separated by commas. If the supplied list is empty, so is the returned string.

### varnum

  Output:       scalar
  Argument:     *varname* (string)

Returns the ID number of the variable called *varname*, or NA is there is no such variable.

### varsimul

  Output:       matrix
  Arguments:    *A* (matrix)
                *U* (matrix)
                *y0* (matrix)

Simulates a $p$-order $n$-variable VAR, that is $y_t = \sum_{i=1}^{p} A_i y_{t-i} + u_t$.

The coefficient matrix $A$ is composed by horizontal stacking of the $A_i$ matrices; it is $n \times np$, with one row per equation. This corresponds to the first $n$ rows of the matrix $compan provided by gretl's `var` and `vecm` commands.

The $u_t$ vectors are contained (as rows) in $U$ ($T \times n$). Initial values are in $y0$ ($p \times n$).

If the VAR contains deterministic terms and/or exogenous regressors, these can be handled by folding them into the $U$ matrix: each row of $U$ then becomes $u_t = B'x_t + e_t$.

The output matrix has $T + p$ rows and $n$ columns; it holds the initial $p$ values of the endogenous variables plus $T$ simulated values.

See also $compan, var, vecm.

### vec

  Output:       column vector
  Argument:   $X$ (matrix)

Stacks the columns of $X$ as a column vector. See also mshape, unvech, vech.

### vech

  Output:       column vector
  Argument:   $A$ (square matrix)

Returns in a column vector the elements of $A$ on and above the diagonal. Typically, this function is used on symmetric matrices; in this case, it can be undone by the function unvech. See also vec.

### weekday

  Output:       scalar
  Arguments:   *year* (scalar)
                *month* (scalar)
                *day* (scalar)

Returns the day of the week (Sunday = 0, Monday = 1, etc.) for the date specified by the three arguments, or `NA` if the date is invalid.

### wmean

  Output:       series
  Arguments:   $Y$ (list)
                $W$ (list)

Returns a series $y$ such that $y_t$ is the weighted mean of the values of the variables in list $Y$ at observation $t$, the respective weights given by the values of the variables in list $W$ at $t$. The weights can therefore be time-varying. The lists $Y$ and $W$ must be of the same length and the weights must be non-negative.

See also wsd, wvar.

### wsd

  Output:       series
  Arguments:   $Y$ (list)
                $W$ (list)

Returns a series $y$ such that $y_t$ is the weighted sample standard deviation of the values of the variables in list $Y$ at observation $t$, the respective weights given by the values of the variables in

list $W$ at $t$. The weights can therefore be time-varying. The lists $Y$ and $W$ must be of the same length and the weights must be non-negative.

See also wmean, wvar.

### wvar

|  |  |
|---|---|
| Output: | series |
| Arguments: | $X$ (list) |
|  | $W$ (list) |

Returns a series $y$ such that $y_t$ is the weighted sample variance of the values of the variables in list $X$ at observation $t$, the respective weights given by the values of the variables in list $W$ at $t$. The weights can therefore be time-varying. The lists $Y$ and $W$ must be of the same length and the weights must be non-negative.

The weighted sample variance is computed as

$$s_w^2 = \frac{n'}{n'-1} \frac{\sum_{i=1}^n w_i (x_i - \bar{x}_w)^2}{\sum_{i=1}^n w_i}$$

where $n'$ is the number of non-zero weights and $\bar{x}_w$ is the weighted mean.

See also wmean, wsd.

### xmax

|  |  |
|---|---|
| Output: | scalar |
| Arguments: | $x$ (scalar) |
|  | $y$ (scalar) |

Returns the greater of $x$ and $y$, or NA if either value is missing.

See also xmin, max, min.

### xmin

|  |  |
|---|---|
| Output: | scalar |
| Arguments: | $x$ (scalar) |
|  | $y$ (scalar) |

Returns the lesser of $x$ and $y$, or NA if either value is missing.

See also xmax, max, min.

### xpx

|  |  |
|---|---|
| Output: | list |
| Argument: | $L$ (list) |

Returns a list that references the squares and cross-products of the variables in list $L$. Squares are named on the pattern sq_*varname* and cross-products on the pattern *var1_var2*. The input variable names are truncated if need be, and the output names may be adjusted in case of duplication of names in the returned list.

### zeromiss

|  |  |
|---|---|
| Output: | same type as input |
| Argument: | $x$ (scalar or series) |

Converts zeros to NAs. If $x$ is a series, the conversion is done element by element. See also missing, misszero, ok.

**zeros**

  Output:      matrix
 Arguments:   $r$ (scalar)
             $c$ (scalar)

Outputs a zero matrix with $r$ rows and $c$ columns. See also ones, seq.

# Chapter 3

# Comments in scripts

When a script does anything non-obvious, it's a good idea to add comments explaining what's going on. This is particularly useful if you plan to share the script with others, but it's also useful as a reminder to yourself — when you revisit a script some months later and wonder what it was supposed to be doing.

The comment mechanism can also be helpful when you're developing a script. There may come a point where you want to execute a script, but bypass execution of some portion of it. Obviously you could delete the portion you wish to bypass, but rather than lose that section you can "comment it out" so that it is ignored by gretl.

Two sorts of comments are supported by gretl. The simpler one is this:

- If a hash mark, #, is encountered in a gretl script, everything from that point to the end of the current line is treated as a comment, and ignored.

If you wish to "comment out" several lines using this mode, you'll have to place a hash mark at the start of each line.

The second sort of comment is patterned after the C programming language:

- If the sequence /* is encountered in a script, all the following input is treated as a comment until the sequence */ is found.

Comments of this sort can extend over several lines. Using this mode it is easy to add lengthy explanatory text, or to get gretl to ignore substantial blocks of commands. As in C, comments of this type cannot be nested.

How do these two comment modes interact? You can think of gretl as starting at the top of a script and trying to decide at each point whether it should or should not be in "ignore mode". In doing so it follows these rules:

- If we're not in ignore mode, then # puts us into ignore mode till the end of the current line.

- If we're not in ignore mode, then /* puts us into ignore mode until */ is found.

This means that each sort of comment can be masked by the other.

- If /* follows # on a given line which does not already start in ignore mode, then there's nothing special about /*, it's just part of a #-style comment.

- If # occurs when we're already in ignore mode, it is just part of a comment.

A few examples follow.

```
/* multi-line comment
   # hello
   # hello */
```

In the above example the hash marks are not special; in particular the hash mark on the third line does not prevent the multi-line comment from terminating at */.

```
# single-line comment /* hello
```

Assuming we were not in ignore mode before the line shown above, it is just a single-line comment: the /* is masked, and does not open a multi-line comment.

You can append a comment to a command:

```
ols 1 0 2 3 # estimate the baseline model
```

Example of "commenting out":

```
/*
# let's skip this for now
ols 1 0 2 3 4
omit 3 4
*/
```

## Chapter 4

# Options, arguments and path-searching

## 4.1 Invoking gretl

`gretl` (under MS Windows, `gretlw32.exe`)[1].

— Opens the program and waits for user input.

`gretl` *datafile*

— Starts the program with the specified datafile in its workspace. The data file may be in any of several formats (see the *Gretl User's Guide*); the program will try to detect the format of the file and treat it appropriately. See also Section 4.4 below for path-searching behavior.

`gretl --help` (or `gretl -h`)

— Print a brief summary of usage and exit.

`gretl --version` (or `gretl -v`)

— Print version identification for the program and exit.

`gretl --english` (or `gretl -e`)

— Force use of English instead of translation.

`gretl --run` *scriptfile* (or `gretl -r` *scriptfile*)

— Start the program and open a window displaying the specified script file, ready to run. See Section 4.4 below for path-searching behavior.

`gretl --db` *database* (or `gretl -d` *database*)

— Start the program and open a window displaying the specified database. If the database files (the `.bin` file and its accompanying `.idx` file) are not in the default system database directory, you must specify the full path. See also the *Gretl User's Guide* for details on databases.

`gretl --dump` (or `gretl -c`)

— Dump the program's configuration information to a plain text file (the name of the file is printed on standard output). May be useful for trouble-shooting.

`gretlw32 --debug` (or `gretlw32 -g`)

— (MS Windows only) Open a console window to display any messages sent to the "standard output" or "standard error" streams. Such messages are not usually visible on Windows; this may be useful for trouble-shooting.

## 4.2 Preferences dialog

Various things in `gretl` are configurable under the "Tools, Preferences" menu. Separate menu items are devoted to the choice of the monospaced font to be used in `gretl` screen output, and, on some platforms, the font used for menus and other messages. The other options are organized under five tabs, as follows.

**General**: Here you can configure the base directory for `gretl`'s shared files. In addition there are several check boxes. Checking "Tell me about gretl updates" makes `gretl` attempt to query the update server at start-up. If your native language setting is not English and the local decimal

---

[1] On Linux, a "wrapper" script named `gretl` is installed. This script checks whether the `DISPLAY` environment variable is set; if so, it launches the GUI program, `gretl_x11`, and if not it launches the command-line program, `gretlcli`

point character is not the period ("."), unchecking "Use locale setting for decimal point" will make gretl use the period regardless. Checking "Allow shell commands" makes it possible to invoke shell commands in scripts and in the gretl console (this facility is disabled by default for security reasons).

**Databases** tab: You can select the directory in which to start looking for native gretl databases; the directory in which to start looking for RATS 4 databases; the host name of the gretl database server to access; and the IP number and port number of the HTTP proxy server to use when contacting the database server (if you're behind a firewall).

**Programs** tab: You can specify the names or paths to various third-party programs that may called by gretl under certain conditions. Note that the item "Command to compile T$_E$X files" can be set to either latex or pdflatex; if latex is selected, T$_E$X output will be previewed in DVI format; if pdflatex is selected, the preview will be in PDF format.

**HCCME** tab: Set preferences regarding robust covariance matrix estimation. See the *Gretl User's Guide* for details.

**Manuals** tab: Select your preferred language for the full gretl documentation in PDF format (currently only English and Italian are supported). When using the English documentation you can also choose between US letter paper and A4 paper.

Settings chosen via the Preferences dialog are stored from one gretl session to the next. Under MS Windows they are stored in the Windows registry; on other platforms they are stored in a plain text file named `.gretl2rc` in the user's home directory.

## 4.3   Invoking gretlcli

`gretlcli`

— Opens the program and waits for user input.

`gretlcli` *datafile*

— Starts the program with the specified datafile in its workspace. The data file may be in any format supported by gretl (see the *Gretl User's Guide* for details). The program will try to detect the format of the file and treat it appropriately. See also Section 4.4 for path-searching behavior.

`gretlcli --help` (or `gretlcli -h`)

— Prints a brief summary of usage.

`gretlcli --version` (or `gretlcli -v`)

— Prints version identification for the program.

`gretlcli --english` (or `gretlcli -e`)

— Force use of English instead of translation.

`gretlcli --run` *scriptfile* (or `gretlcli -r` *scriptfile*)

— Execute the commands in *scriptfile* then hand over input to the command line. See Section 4.4 for path-searching behavior.

`gretlcli --batch` *scriptfile* (or `gretlcli -b` *scriptfile*)

— Execute the commands in *scriptfile* then exit. When using this option you will probably want to redirect output to a file. See Section 4.4 for path-searching behavior.

When using the `--run` and `--batch` options, the script file in question must call for a data file to be opened. This can be done using the `open` command within the script.

## 4.4   Path searching

When the name of a data file or script file is supplied to gretl or gretlcli on the command line, the file is looked for as follows:

1. "As is". That is, in the current working directory or, if a full path is specified, at the specified location.

2. In the user's gretl directory (see Table 4.1 for the default values; note that PERSONAL is a placeholder that is expanded by Windows in a user- and language-specific way, typically involving "My Documents" on English-language systems).

3. In any immediate sub-directory of the user's gretl directory.

4. In the case of a data file, search continues with the main gretl data directory. In the case of a script file, the search proceeds to the system script directory. See Table 4.1 for the default settings. (PREFIX denotes the base directory chosen at the time gretl is installed.)

5. In the case of data files the search then proceeds to all immediate sub-directories of the main data directory.

**Table 4.1**: Default path settings

|                          | *Linux*                     | *MS Windows*           |
| ------------------------ | --------------------------- | ---------------------- |
| User directory           | `$HOME/gretl`               | `PERSONAL\gretl`       |
| System data directory    | `PREFIX/share/gretl/data`   | `PREFIX\gretl\data`    |
| System script directory  | `PREFIX/share/gretl/scripts`| `PREFIX\gretl\scripts` |

Thus it is not necessary to specify the full path for a data or script file unless you wish to override the automatic searching mechanism. (This also applies within gretlcli, when you supply a filename as an argument to the open or run commands.)

When a command script contains an instruction to open a data file, the search order for the data file is as stated above, except that the directory containing the script is also searched, immediately after trying to find the data file "as is".

**MS Windows**

Under MS Windows configuration information for gretl and gretlcli is stored in the Windows registry. A suitable set of registry entries is created when gretl is first installed, and the settings can be changed under gretl's "Tools, Preferences" menu. In case anyone needs to make manual adjustments to this information, the entries can be found (using the standard Windows program regedit.exe) under Software\gretl in HKEY_LOCAL_MACHINE (the main gretl directory and the paths to various auxiliary programs) and HKEY_CURRENT_USER (all other configurable variables).

# Chapter 5

# Reserved Words

Reserved words, which cannot be used as the names of variables, fall into the following categories:

- Names of constants and data types, plus a few specials: `const`, `CONST`, `pi`, `NA`, `null`, `obs`, `scalar`, `series`, `matrix`, `string`, `list`, `bundle`, `kalman`, `void`, `continue`, `next`, `to`.

- Names of gretl commands (see section 1.2).

User-defined functions cannot have names which collide with built-in functions, the names of which are shown in Table 5.1.

**Table 5.1**: Function names

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| BFGSmax | I | NRmax | abs | acos | acosh | argname | asin |
| asinh | atan | atanh | bessel | bkfilt | boxcox | bwfilt | cdemean |
| cdf | cdiv | ceil | cholesky | chowlin | cmult | cnorm | colname |
| colnames | cols | corr | corrgm | cos | cosh | cov | critical |
| cum | deseas | det | diag | diagcat | diff | digamma | dnorm |
| dsort | dummify | eigengen | eigensym | eigsolve | epochday | errmsg | exp |
| fcstats | fdjac | fft | ffti | filter | firstobs | floor | fracdiff |
| fraclag | freq | gammafun | getenv | gini | ginv | grab | hdprod |
| hpfilt | imaxc | imaxr | imhof | iminc | iminr | inbundle | infnorm |
| inlist | int | inv | invcdf | invmills | invpd | irf | irr |
| isconst | islist | isnull | isscalar | isseries | isstring | kdensity | kfilter |
| ksimul | ksmooth | kurtosis | lags | lastobs | ldet | ldiff | lincomb |
| ljungbox | ln | lngamma | loess | log | log10 | log2 | logistic |
| lower | lrvar | max | maxc | maxr | mcorr | mcov | mcovg |
| mean | meanc | meanr | median | mexp | min | minc | minr |
| missing | misszero | mlag | mnormal | mols | monthlen | movavg | mpols |
| mrandgen | mread | mreverse | mrls | mshape | msortby | muniform | mwrite |
| mxtab | nadarwat | nelem | ngetenv | nobs | normal | npv | nullspace |
| obslabel | obsnum | ok | onenorm | ones | orthdev | pdf | pergm |
| pmax | pmean | pmin | pnobs | polroots | polyfit | princomp | psd |
| psdroot | pshrink | pvalue | pxsum | qform | qnorm | qrdecomp | quantile |
| randgen | randint | rank | ranking | rcond | readfile | replace | resample |
| round | rownames | rows | sd | sdc | sdiff | selifc | selifr |
| seq | setnote | simann | sin | sinh | skewness | sort | sortby |
| sqrt | sscanf | sst | strlen | strncmp | strsplit | strstr | strsub |
| sum | sumall | sumc | sumr | svd | tan | tanh | toepsolv |
| tolower | tr | transp | trimr | uniform | uniq | unvech | upper |
| urcpval | values | var | varname | varnum | varsimul | vec | vech |
| weekday | wmean | wsd | wvar | xmax | xmin | xpx | zeromiss |
| zeros | | | | | | | |

# Bibliography

Agresti, A. (1992) 'A survey of exact inference for contingency tables', *Statistical Science* 7: 131–153.

Akaike, H. (1974) 'A new look at the statistical model identification', *IEEE Transactions on Automatic Control* AC-19: 716–723.

Arellano, M. and S. Bond (1991) 'Some tests of specification for panel data: Monte carlo evidence and an application to employment equations', *The Review of Economic Studies* 58: 277–297.

Belsley, D., E. Kuh and R. Welsch (1980) *Regression Diagnostics*, New York: Wiley.

Choi, I. (2001) 'Unit root tests for panel data', *Journal of International Money and Finance* 20(2): 249–272.

Chow, G. C. and A.-l. Lin (1971) 'Best linear unbiased interpolation, distribution, and extrapolation of time series by related series', *The Review of Economics and Statistics* 53(4): 372–375.

Cleveland, W. S. (1979) 'Robust locally weighted regression and smoothing scatterplots', *Journal of the American Statistical Association* 74(368): 829–836.

Davidson, R. and J. G. MacKinnon (1993) *Estimation and Inference in Econometrics*, New York: Oxford University Press.

———— (2004) *Econometric Theory and Methods*, New York: Oxford University Press.

Doornik, J. A. (1998) 'Approximations to the asymptotic distribution of cointegration tests', *Journal of Economic Surveys* 12: 573–593. Reprinted with corrections in McAleer and Oxley (1999).

Edgerton, D. and C. Wells (1994) 'Critical values for the cusumsq statistic in medium and large sized samples', *Oxford Bulletin of Economics and Statistics* 56: 355–365.

Elliott, G., T. J. Rothenberg and J. H. Stock (1996) 'Efficient tests for an autoregressive unit root', *Econometrica* 64: 813–836.

Engle, R. F. and C. W. J. Granger (1987) 'Co-integration and error correction: Representation, estimation, and testing', *Econometrica* 55: 251–276.

Fiorentini, G., G. Calzolari and L. Panattoni (1996) 'Analytic derivatives and the computation of GARCH estimates', *Journal of Applied Econometrics* 11: 399–417.

Geweke, J. and S. Porter-Hudak (1983) 'The estimation and application of long memory time series models', *Journal of Time Series Analysis* 4: 221–238.

Godfrey, L. G. (1994) 'Testing for serial correlation by variable addition in dynamic models estimated by instrumental variables', *The Review of Economics and Statistics* 76(3): 550–559.

Golub, G. H. and C. F. Van Loan (1996) *Matrix Computations*, Baltimore and London: The John Hopkins University Press, third edn.

Greene, W. H. (2000) *Econometric Analysis*, Upper Saddle River, NJ: Prentice-Hall, fourth edn.

Hamilton, J. D. (1994) *Time Series Analysis*, Princeton, NJ: Princeton University Press.

Heckman, J. (1979) 'Sample selection bias as a specification error', *Econometrica* 47: 153–161.

Im, K. S., M. H. Pesaran and Y. Shin (2003) 'Testing for unit roots in heterogeneous panels', *Journal of Econometrics* 115: 53–74.

Imhof, J. P. (1961) 'Computing the distribution of quadratic forms in normal variables', *Biometrika* 48: 419–426.

Kiviet, J. F. (1986) 'On the rigour of some misspecification tests for modelling dynamic relationships', *Review of Economic Studies* 53: 241–261.

Koenker, R. (1981) 'A note on studentizing a test for heteroscedasticity', *Journal of Econometrics* 17: 107–112.

———— (1994) 'Confidence intervals for regression quantiles'. In P. Mandl and M. Huskova (eds.), *Asymptotic Statistics*, pp. 349–359. New York: Springer-Verlag.

Koenker, R. and G. Bassett (1978) 'Regression quantiles', *Econometrica* 46: 33–50.

Koenker, R. and J. Machado (1999) 'Goodness of fit and related inference processes for quantile regression', *Journal of the American Statistical Association* 94: 1296–1310.

Koenker, R. and Q. Zhao (1994) 'L-estimation for linear heteroscedastic models', *Journal of Nonparametric Statistics* 3: 223–235.

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt and Y. Shin (1992) 'Testing the null of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?', *Journal of Econometrics* 54: 159–178.

Levin, A., C.-F. Lin and J. Chu (2002) 'Unit root tests in panel data: asymptotic and finite-sample properties', *Journal of Econometrics* 108: 1–24.

Locke, C. (1976) 'A test for the composite hypothesis that a population has a gamma distribution', *Communications in Statistics — Theory and Methods* A5: 351–364.

MacKinnon, J. G. (1996) 'Numerical distribution functions for unit root and cointegration tests', *Journal of Applied Econometrics* 11: 601–618.

Maddala, G. S. (1992) *Introduction to Econometrics*, Englewood Cliffs, NJ: Prentice-Hall.

Marsaglia, G. and W. W. Tsang (2000) 'The ziggurat method for generating random variables', *Journal of Statistical Software* 5: 3–30.

McAleer, M. and L. Oxley (1999) *Practical Issues in Cointegration Analysis*, Oxford: Blackwell.

Nerlove, M. (1971) 'Further evidence on the estimation of dynamic economic relations from a time series of cross sections', *Econometrica* 39: 359–382.

Neter, J., W. Wasserman and M. H. Kutner (1990) *Applied Linear Statistical Models*, Boston: Irwin, third edn.

Pesaran, M. H. and L. W. Taylor (1999) 'Diagnostics for IV regressions', *Oxford Bulletin of Economics and Statistics* 61(2): 255–281.

Phillips, P. C. B. and K. Shimotsu (2004) 'Local Whittle estimation in nonstationary and unit root cases', *The Annals of Statistics* 32(2): 659–692. http://arxiv.org/pdf/math/0406462.

Ramanathan, R. (2002) *Introductory Econometrics with Applications*, Fort Worth: Harcourt, fifth edn.

Robinson, P. (1995) 'Gaussian semiparametric estimation of long range dependence', *Annals of Statistics* 22: 1630–1661.

Saito, M. and M. Matsumoto (2008) 'SIMD-oriented Fast Mersenne Twister: a 128-bit pseudorandom number generator'. In A. Keller, S. Heinrich and H. Niederreiter (eds.), *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pp. 607–622. Berlin: Springer.

Schwert, G. W. (1989) 'Tests for unit roots: A Monte Carlo investigation', *Journal of Business and Economic Statistics* 7(2): 5–17.

Sephton, P. S. (1995) 'Response surface estimates of the KPSS stationarity test', *Economics Letters* 47: 255–261.

Shapiro, S. and L. Chen (2001) 'Composite tests for the gamma distribution', *Journal of Quality Technology* 33: 47–59.

Stock, J. H. and M. W. Watson (2003) *Introduction to Econometrics*, Boston: Addison-Wesley.

Stock, J. H., J. H. Wright and M. Yogo (2002) 'A survey of weak instruments and weak identification in generalized method of moments', *Journal of Business & Economic Statistics* 20(4): 518–529.

Stock, J. H. and M. Yogo (2003) 'Testing for weak instruments in linear IV regression'. NBER Technical Working Paper 284. http://www.nber.org/~myogo/papers/published/RothenbergCh5.pdf.

Swamy, P. A. V. B. and S. S. Arora (1972) 'The exact finite sample properties of the estimators of coefficients in the error components regression models', *Econometrica* 40: 261–275.

Windmeijer, F. (2005) 'A finite sample correction for the variance of linear efficient two-step GMM estimators', *Journal of Econometrics* 126: 25–51.