

## Saved Results

`dfgls` saves the following scalars in `r()`:

<code>r(N)</code>	number of observations
<code>r(optlag)</code>	optimal lag order
<code>r(sc<math>n</math>)</code>	Schwarz criterion at lag $n$
<code>r(rmse<math>n</math>)</code>	root mean square error at lag $n$
<code>r(dft<math>n</math>)</code>	DF-GLS statistic at lag $n$

`kpss` saves the following scalars in `r()`:

<code>r(N)</code>	number of observations
<code>r(dft<math>n</math>)</code>	KPSS statistic at lag $n$

## Acknowledgments

I acknowledge useful conversations with Serena Ng, James Stock, and Vince Wiggins. The KPSS code was adapted from John Barkoulas' RATS code for that test. Thanks also to Richard Sperling for tracking down a discrepancy between published work and the `dfgls` output and alerting me to the Cheung and Lai estimates. Any remaining errors are my own.

## References

- Cheung, Y. W. and K.-S. Lai. 1995. Lag order and critical values of a modified Dickey–Fuller test. *Oxford Bulletin of Economics and Statistics* 57: 411–419.
- Elliott, G., T. J. Rothenberg, and J. H. Stock. 1996. Efficient tests for an autoregressive unit root. *Econometrica* 64: 813–836.
- Kwiatkowski, D., P. C. Phillips, P. Schmidt, and Y. Shin. 1992. Testing the null hypothesis 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.
- Lee, D. and P. Schmidt. 1996. On the power of the KPSS test of stationarity against fractionally-integrated alternatives. *Journal of Econometrics* 73: 285–302.
- Ng, S. and P. Perron. 1995. Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association* 90: 268–281.
- Schwert, G. W. 1989. Tests for unit roots: A Monte Carlo investigation. *Journal of Business and Economic Statistics* 7: 147–160.
- Stock, J. H. 1994. Unit roots, structural breaks and trends. In *Handbook of Econometrics IV*, ed. R. F. Engle and D. L. McFadden. Amsterdam: Elsevier.

sts16	Tests for long memory in a time series
-------	--

Christopher F. Baum, Boston College, baum@bc.edu  
Vince Wiggins, Stata Corporation, vwiggins@stata.com

**Abstract:** Implements the Geweke/Porter-Hudak log periodogram estimator (1983), the Phillips modified log periodogram estimator (1999b) and the Robinson log periodogram estimator (1995) for the diagnosis of long memory, or fractional integration, in a time series. The Robinson estimator may be applied to a set of time series.

**Keywords:** fractional integration, long memory, stationarity, time series.

## Syntax

```
gphudak varname [if exp] [in range] [, powers(numlist) ]
```

```
modlpr varname [if exp] [in range] [, powers(numlist) notrend ]
```

```
roblpr varlist [if exp] [in range] [, powers(numlist) l(#) j(#) constraints(numlist) ]
```

These tests are for use with time series data; you must `tsset` your data before using these tests; see [R] `tsset`. `varname` or `varlist` may contain time series operators; see [U] **14.4.3 Time-series varlists**.

## Options

`powers(numlist)` indirectly specifies the number of ordinates to be included in the regression. A number of ordinates equal to the integer part of  $T$  raised to the `powers(numlist)` will be used. Powers ranging from 0.50 to 0.75 are commonly employed for `gphudak` and `modlpr`. These routines use the default power of 0.5. `roblpr` uses the default power of 0.9. For `roblpr`, multiple powers may only be specified if a single variable appears in `varlist`.

`notrend` specifies that detrending is not to be applied by `modlpr`. By default, a linear trend will be removed from the series.

1 (#) specifies the number of initial ordinates to be removed from the regression for `rob1pr`. Some researchers have found that such exclusion improves the properties of tests based on log-periodogram regressions. The default value of 1 is zero.

j (#) specifies that the log periodogram employed in `rob1pr` is to be computed as an average of adjacent ordinates. The default value of j is 1, so that no averaging is performed. If j is 2, the number of ordinates is halved; with a j of 3, divided by three, and so on. When j is greater than 1, the value of `powers` should be set large enough so that the averaged ordinates are sufficient in number.

`constraints(numlist)` specifies the constraint numbers of the linear constraints to be applied during estimation in `rob1pr`. The default is to perform unconstrained estimation. This option allows the imposition of linear constraints prior to estimation of the pooled coefficient vector. For instance, if `varlist` contains prices, dividends, and returns, and your prior (or previous findings) states that prices' and dividends' order of integration is indistinguishable, one might impose that constraint to improve the power of the  $F$  test provided by `rob1pr`. You would specify the constraints prior to the `rob1pr` command and then provide the list of constraints in the `constraints` option to `rob1pr`.

*Technical note on constraints.* When constraints are imposed it is difficult to identify the number of numerator degrees of freedom in the test for equality of  $d$  coefficients reported at the bottom of `rob1pr`'s output. Since constraints can be of any general form and it is possible to specify constraints that are not unique, `rob1pr` determines the degrees of freedom from the rank of the matrix used to compute the Wald statistic. Determining that matrix rank from a numerical standpoint can be problematic, in which case `rob1pr` may overstate the number of constraints being tested and thereby incorrectly compute the numerator degrees of freedom for the test. This rarely has a meaningful impact on the statistical test, but you may wish to test only the unconstrained coefficients if the computed degrees of freedom are wrong.

For example, after the final example below, we could perform the test by typing `test ftap == ftaret`. In this case, the degrees of freedom were correct, so we needn't have gone to the trouble.

## Description

The model of an autoregressive fractionally integrated moving average process of a time series of order  $(p, d, q)$ , denoted by ARFIMA( $p, d, q$ ), with mean  $\mu$ , may be written using operator notation as

$$\Phi(L)(1-L)^d(y_t - \mu) = \Theta(L)\epsilon_t, \quad \epsilon_t \sim i.i.d.(0, \sigma_\epsilon^2) \quad (1)$$

where  $L$  is the backward-shift operator,

$$\Phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$$

$\Theta(L) = 1 + \vartheta_1 L + \dots + \vartheta_q L^q$ , and  $(1-L)^d$  is the fractional differencing operator defined by

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)L^k}{\Gamma(-d)\Gamma(k+1)} \quad (2)$$

with  $\Gamma(\cdot)$  denoting the gamma (generalized factorial) function. The parameter  $d$  is allowed to assume any real value. The arbitrary restriction of  $d$  to integer values gives rise to the standard autoregressive integrated moving average (ARIMA) model. The stochastic process  $y_t$  is both stationary and invertible if all roots of  $\Phi(L)$  and  $\Theta(L)$  lie outside the unit circle and  $|d| < 0.5$ . The process is nonstationary for  $d \geq 0.5$ , as it possesses infinite variance; for example, see Granger and Joyeux (1980).

Assuming that  $d \in [0, 0.5)$ , Hosking (1981) showed that the autocorrelation function,  $\rho(\cdot)$ , of an ARFIMA process is proportional to  $k^{2d-1}$  as  $k \rightarrow \infty$ . Consequently, the autocorrelations of the ARFIMA process decay hyperbolically to zero as  $k \rightarrow \infty$  in contrast to the faster, geometric decay of a stationary ARMA process. For  $d \in (0, 0.5)$ ,  $\sum_{j=-n}^n |\rho(j)|$  diverges as  $n \rightarrow \infty$ , and the ARFIMA process is said to exhibit long memory, or long-range positive dependence. The process is said to exhibit intermediate memory (anti-persistence), or long-range negative dependence, for  $d \in (-0.5, 0)$ . The process exhibits short memory for  $d = 0$ , corresponding to stationary and invertible ARMA modeling. For  $d \in [0.5, 1)$  the process is mean reverting, even though it is not covariance stationary, as there is no long-run impact of an innovation on future values of the process.

If a series exhibits long memory, it is neither stationary ( $I(0)$ ) nor is it a unit root ( $I(1)$ ) process; it is an  $I(d)$  process, with  $d$  a real number. A series exhibiting long memory, or persistence, has an autocorrelation function that damps hyperbolically, more slowly than the geometric damping exhibited by "short memory" (ARMA) processes. Thus, it may be predictable at long horizons. Long memory models originated in hydrology and have been widely applied in economics and finance. An excellent survey of long memory models is given by Baillie (1996).

There are two approaches to the estimation of an ARFIMA  $(p, d, q)$  model: exact maximum likelihood estimation, as proposed by Sowell (1992), and semiparametric approaches, as described in this insert. Sowell's approach requires specification of the  $p$  and  $q$  values, and estimation of the full ARFIMA model conditional on those choices. This involves all the attendant

difficulties of choosing an appropriate ARMA specification, as well as a formidable computational task for each combination of  $p$  and  $q$  to be evaluated. The methods described here assume that the short memory or ARMA components of the time series are relatively unimportant, so that the long memory parameter  $d$  may be estimated without fully specifying the data-generating process. These methods are thus described as semiparametric.

`gphudak` performs the Geweke and Porter-Hudak (GPH 1983) semiparametric log periodogram regression, often described as the “GPH test,” for long memory (fractional integration) in a time series. The GPH method uses nonparametric methods—a spectral regression estimator—to evaluate  $d$  without explicit specification of the ARMA parameters of the series. The series is usually differenced so that the resulting  $d$  estimate will fall in the  $[-0.5, 0.5]$  interval.

Geweke and Porter-Hudak (1983) proposed a semiparametric procedure to obtain an estimate of the memory parameter  $d$  of a fractionally integrated process  $X_t$  in a model of the form

$$(1 - L)^d X_t = \epsilon_t, \quad (3)$$

where  $\epsilon_t$  is stationary with zero mean and continuous spectral density  $f_\epsilon(\lambda) > 0$ . The estimate  $\hat{d}$  is obtained from the application of ordinary least squares to

$$\log(I_x(\lambda_s)) = \hat{c} - \hat{d} \log|1 - e^{i\lambda_s}|^2 + \text{residual} \quad (4)$$

computed over the fundamental frequencies  $\{\lambda_s = 2\pi s/n, s = 1, \dots, m < n\}$ . We define

$$\omega_x(\lambda_s) = \frac{1}{\sqrt{2\pi n}} \sum_{t=1}^n X_t e^{it\lambda_s}$$

as the discrete Fourier transform (DFT) of the time series  $X_t$ ,  $I_x(\lambda_s) = \omega_x(\lambda_s) \omega_x(\lambda_s)^*$  as the periodogram, and  $x_s = \log|1 - e^{i\lambda_s}|$ . Ordinary least squares on (4) yields

$$\hat{d} = \frac{\sum_{s=1}^m x_s \log I_x(\lambda_s)}{2 \sum_{s=1}^m x_s^2} \quad (5)$$

Various authors have proposed methods for the choice of  $m$ , the number of Fourier frequencies included in the regression. The regression slope estimate is an estimate of the slope of the series’ power spectrum in the vicinity of the zero frequency; if too few ordinates are included, the slope is calculated from a small sample. If too many are included, medium and high-frequency components of the spectrum will contaminate the estimate. A choice of  $\sqrt{T}$  or 0.5 for `power` is often employed. To evaluate the robustness of the GPH estimate, a range of power values (from 0.40 to 0.75) is commonly calculated as well. Two estimates of the  $d$  coefficient’s standard error are commonly employed: the regression standard error, giving rise to a standard  $t$  test, and an asymptotic standard error, based upon the theoretical variance of the log periodogram of  $\pi^2/6$ . The statistic based upon that standard error has a standard normal distribution under the null.

`modlpr` computes a modified form of the GPH estimate of the long memory parameter,  $d$ , of a time series, proposed by Phillips (1999a, 1999b). Phillips (1999a) points out that the prior literature on this semiparametric approach does not address the case of  $d = 1$ , or a unit root, in (3), despite the broad interest in determining whether a series exhibits unit-root behavior or long memory behavior, and his work showing that the  $\hat{d}$  estimate of (5) is inconsistent when  $d > 1$ , with  $\hat{d}$  exhibiting asymptotic bias toward unity. This weakness of the GPH estimator is solved by Phillips’ modified log periodogram regression estimator, in which the dependent variable is modified to reflect the distribution of  $d$  under the null hypothesis that  $d = 1$ . The estimator gives rise to a test statistic for  $d = 1$  which is a standard normal variate under the null. Phillips suggests that deterministic trends should be removed from the series before application of the estimator. Accordingly, the routine will automatically remove a linear trend from the series. This may be suppressed with the `notrend` option. The comments above regarding `power` apply equally to `modlpr`.

The Phillips (1999b) modification of the GPH estimator is based on an exact representation of the DFT in the unit root case. The modification expresses

$$\omega_x(\lambda_s) = \frac{\omega_u(\lambda_s)}{1 - e^{i\lambda_s}} - \frac{e^{i\lambda_s}}{1 - e^{i\lambda_s}} \frac{X_n}{\sqrt{2\pi n}}$$

and the modified DFT as

$$v_x(\lambda_s) = \omega_x(\lambda_s) + \frac{e^{i\lambda_s}}{1 - e^{i\lambda_s}} \frac{X_n}{\sqrt{2\pi n}}$$

with associated periodogram ordinates  $I_v(\lambda_s) = v_x(\lambda_s) v_x(\lambda_s)^*$  (Phillips 1999b, 9). He notes that both  $v_x(\lambda_s)$  and, thus,  $I_v(\lambda_s)$  are observable functions of the data. The log-periodogram regression is now the regression of  $\log I_v(\lambda_s)$  on  $a_s = \log |1 - e^{i\lambda_s}|$ . Defining  $\bar{a} = m^{-1} \sum_{s=1}^m a_s$  and  $x_s = a_s - \bar{a}$ , the modified estimate of the long-memory parameter becomes

$$\tilde{d} = \frac{\sum_{s=1}^m x_s \log I_v(\lambda_s)}{2 \sum_{s=1}^m x_s^2} \quad (6)$$

Phillips proves that, with appropriate assumptions on the distribution of  $\epsilon_t$ , the distribution of  $\tilde{d}$  follows

$$\sqrt{m}(\tilde{d} - d) \rightarrow N\left(0, \frac{\pi^2}{24}\right) \quad (7)$$

in distribution, so  $\tilde{d}$  has the same limiting distribution at  $d = 1$  as does the GPH estimator in the stationary case so  $\tilde{d}$  is consistent for values of  $d$  around unity. A semiparametric test statistic for a unit root against a fractional alternative is then based upon the statistic (Phillips 1999a, 10)

$$z_d = \frac{\sqrt{m}(\tilde{d} - 1)}{\pi/24} \quad (8)$$

with critical values from the standard normal distribution. This test is consistent against both  $d < 1$  and  $d > 1$  fractional alternatives.

`rob1pr` computes the Robinson (1995) multivariate semiparametric estimate of the long memory (fractional integration) parameters,  $d(g)$ , of a set of  $G$  time series,  $y(g)$ ,  $g = 1, \dots, G$  with  $G \geq 1$ . When applied to a set of time series, the  $d(g)$  parameter for each series is estimated from a single log-periodogram regression which allows the intercept and slope to differ for each series. One of the innovations of Robinson's estimator is that it is not restricted to using a small fraction of the ordinates of the empirical periodogram of the series, that is, the reasonable values of `power` need not exclude a sizable fraction of the original sample size. The estimator also allows for the removal of one or more initial ordinates and for the averaging of the periodogram over adjacent frequencies. The rationale for using non-default values of either of these options is presented in Robinson (1995).

Robinson (1995) proposes an alternative log-periodogram regression estimator which he claims provides "modestly superior asymptotic efficiency to  $\bar{d}(0)$ ", ( $\bar{d}(0)$  being the Geweke and Porter-Hudak estimator) Robinson (1995, 1052). Robinson's formulation of the log-periodogram regression also allows for the formulation of a multivariate model, providing justification for tests that different time series share a common differencing parameter. Normality of the underlying time series is assumed, but Robinson claims that other conditions underlying his derivation are milder than those conjectured by GPH.

We present here Robinson's multivariate formulation, which applies to a single time series as well. Let  $X_t$  represent a  $G$ -dimensional vector with  $g^{th}$  element  $X_{gt}$ ,  $g = 1, \dots, G$ . Assume that  $X_t$  has a spectral density matrix  $\int_{-\pi}^{\pi} e^{ij\lambda} f(\lambda) d\lambda$ , with  $(g, h)$  element denoted as  $f_{gh}(\lambda)$ . The  $g$ th diagonal element,  $f_{gg}(\lambda)$ , is the power spectral density of  $X_{gt}$ . For  $0 < C_g < \infty$  and  $-1/2 < d_g < 1/2$ , assume that  $f_{gg}(\lambda) \sim C_g \lambda^{-2d_g}$  as  $\lambda \rightarrow 0+$  for  $g = 1, \dots, G$ . The periodogram of  $X_{gt}$  is then denoted as

$$I_g(\lambda) = (2\pi n)^{-1} \left| \sum_{t=1}^n X_{gt} e^{it\lambda} \right|^2, g = 1, \dots, G \quad (9)$$

Without averaging the periodogram over adjacent frequencies nor omission of  $l$  initial frequencies from the regression, we may define  $Y_{gk} = \log I_g(\lambda_k)$ . The least squares estimates of  $c = (c_1, \dots, c_G)'$  and  $d = (d_1, \dots, d_G)'$  are given by

$$\begin{bmatrix} \tilde{c} \\ \tilde{d} \end{bmatrix} = \text{vec} \{ Y' Z (Z' Z)^{-1} \} \quad (10)$$

where  $Z = (Z_1, \dots, Z_m)'$ ,  $Z_k = (1, -2 \log \lambda_k)'$ ,  $Y = (Y_1, \dots, Y_G)$ , and  $Y_g = (Y_{g,1}, \dots, Y_{g,m})'$  for  $m$  periodogram ordinates. Standard errors for  $\tilde{d}_g$  and for a test of the restriction that two or more of the  $d_g$  are equal may be derived from the estimated covariance matrix of the least squares coefficients. The standard errors for the estimated parameters are derived from a pooled estimate of the variance in the multivariate case, so that their interval estimates differ from those of their univariate counterparts. Modifications to this derivation when the frequency-averaging (j) or omission of initial frequencies (l) options are selected may be found in Robinson (1995).

## Examples

Data from Terence Mills' *Econometric Analysis of Financial Time Series* on UK FTA All Share stock returns (`ftaret`) and dividends (`ftadiv`) are analyzed.

```

. use http://fmwww.bc.edu/ec-p/data/Mills2d/fta.dta
. tsset
      time variable:  month, 1965m1 to 1995m12
. gphudak ftaret,power(0.5 0.6 0.7)
GPH estimate of fractional differencing parameter
-----
Power   Ords   Est d   StdErr  t(H0: d=0)  P>|t|   Asy.
        |         |         |         |         |         |         |
StdErr  z(H0: d=0)  P>|z|
-----+-----
.50     20   -.00204  .160313  -0.0127   0.990   .187454
.60     35   .228244  .145891   1.5645   0.128   .130206
.70     64   .141861  .089922   1.5776   0.120   .091267
        |         |         |         |         |         |         |
-----+-----

. modlpr ftaret, power(0.5 0.55:0.8)
Modified LPR estimate of fractional differencing parameter
-----
Power   Ords   Est d   Std Err  t(H0: d=0)  P>|t|   z(H0: d=1)  P>|z|
-----+-----
.50     19   .0231191  .139872   0.1653   0.870   -6.6401
.55     25   .2519889  .1629533  1.5464   0.135   -5.8322
.60     34   .2450011  .1359888  1.8016   0.080   -6.8650
.65     46   .1024504  .1071614  0.9560   0.344   -9.4928
.70     63   .1601207  .0854082  1.8748   0.065   -10.3954
.75     84   .1749659  .081113   2.1566   0.034   -11.7915
.80    113   .0969439  .0676039  1.4340   0.154   -14.9696
        |         |         |         |         |         |         |
-----+-----

. roblpr ftaret
Robinson estimates of fractional differencing parameter
-----
Power   Ords   Est d   Std Err  t(H0: d=0)  P>|t|
-----+-----
.90    205   .1253645  .0446745   2.8062   0.005
        |         |         |         |         |
-----+-----

. roblpr ftap ftadiv
Robinson estimates of fractional differencing parameters
Power = .90                               Ords = 205
-----+-----
Variable   |   Est d   Std Err   t   P>|t|
-----+-----
ftap      |   .8698092  .0163302  53.2640  0.000
ftadiv    |   .8717427  .0163302  53.3824  0.000
-----+-----

Test for equality of d coefficients:  F(1,406) = .00701  Prob > F = 0.9333

. constraint define 1 ftap=ftadiv
. roblpr ftap ftadiv ftaret, c(1)
Robinson estimates of fractional differencing parameters
Power = .90                               Ords = 205
-----+-----
Variable   |   Est d   Std Err   t   P>|t|
-----+-----
ftap      |   .8707759  .0205143  42.4473  0.000
ftadiv    |   .8707759  .0205143  42.4473  0.000
ftaret    |   .1253645  .0290116   4.3212  0.000
-----+-----

Test for equality of d coefficients:  F(1,610) = 440.11  Prob > F = 0.0000

```

The GPH test, applied to the stock returns series, generates estimates of the long memory parameter that cannot reject the null at the ten percent level using the  $t$  test. Phillips' modified LPR, applied to this series, finds that  $d = 1$  can be rejected for all powers tested, while  $d = 0$  (stationarity) may be rejected at the ten percent level for powers 0.6, 0.7, and 0.75. Robinson's estimate for the returns series alone is quite precise. Robinson's multivariate test, applied to the price and dividends series, finds that each series has  $d > 0$ . The test that they share the same  $d$  cannot be rejected. Accordingly, the test is applied to all three series subject to the constraint that price and dividends series have a common  $d$ , yielding a more precise estimate of the difference in  $d$  parameters between those series and the stock returns series.

## Saved Results

`gphudak` saves in `e()`:

<code>e(N_powers)</code>	number of powers (scalar)
<code>e(depvar)</code>	dependent variable name (macro)
<code>e(gph)</code>	matrix of results, 9 by <code>N_powers</code>

`modlpr` saves in `e()`:

<code>e(N_powers)</code>	number of powers (scalar)
<code>e(depvar)</code>	dependent variable name (macro)
<code>e(modlpr)</code>	matrix of results, 8 by <code>N_powers</code>

`roblpr` saves the following scalars in `r()`:

<code>r(N)</code>	number of observations
<code>r(rob)</code>	$d$ estimate
<code>r(se)</code>	estimated standard error of $d$
<code>r(t)</code>	$t$ statistic
<code>r(p)</code>	$p$ -value of $t$ statistic

If more than one power is specified in `roblpr`, the saved results pertain to the last power used.

## Acknowledgments

The first author acknowledges John Barkoulas' original exposition of the ARFIMA model, and thanks Peter Phillips for clarifying comments on his working papers. Any remaining errors are the authors' responsibility.

## References

- Baillie, R. 1996. Long memory processes and fractional integration in econometrics. *Journal of Econometrics* 73: 5–59.
- 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.
- Granger, C. W. J. and R. Joyeux. 1980. An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1: 15–39.
- Hosking, J. R. M. 1981. Fractional differencing. *Biometrika* 68: 165–176.
- Phillips, P. C. B. 1999a. Discrete Fourier transforms of fractional processes. Unpublished working paper No. 1243, Cowles Foundation for Research in Economics, Yale University. <http://cowles.econ.yale.edu/P/cd/d12a/d1243.pdf>
- . 1999b. Unit root log periodogram regression. Unpublished working paper No. 1244, Cowles Foundation for Research in Economics, Yale University. <http://cowles.econ.yale.edu/P/cd/d12a/d1244.pdf>
- Robinson, P. M. 1995. Log-periodogram regression of time series with long range dependence. *Annals of Statistics* 23: 1048–1072.
- Sowell, F. 1992. Maximum likelihood estimation of stationary univariate fractionally-integrated time-series models, *Journal of Econometrics* 53: 165–188.

sts17	Compacting time series data
-------	-----------------------------

Christopher F. Baum, Boston College, [baum@bc.edu](mailto:baum@bc.edu)

**Abstract:** `tscollap` provides the ability to compact data of monthly, quarterly or half-yearly frequencies to a lower frequency by one or more methods (e.g., average, sum, last value per period, and so on).

**Keywords:** time series, data frequency, collapse.

## Syntax

`tscollap` *clist*, `to`(*freq*) [generate(*freqvar*) ]

where *clist* is either

[ (*stat*) ] *varlist* [ [ (*stat*) ] ... ]

or

[ (*stat*) *target\_var* = *varname* [*target\_var* = *varname* ... ] [ [ (*stat*) ... ] ]

or any combination of the *varlist* or *target\_var* forms, and *stat* is one of