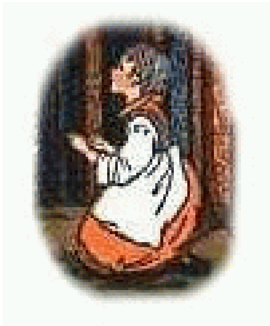
Manual dos Comandos do Gretl



Gnu Regression, Econometrics and Time-series Library

Allin Cottrell
Department of Economics
Wake Forest university

Hélio Guilherme
Tradução Portuguesa

Maio de 2007

Conteúdo

add

Argumento:	*lista-de-variáveis*
Opções:	`--vcv` (mostrar matriz de covariância)
	`--quiet` (não mostrar estimativas para o modelo aumentado)
	`--silent` (não mostrar nada)
	`--inst` (acrescentar como instrumento, apenas para TSLS)
	`--both` (acrescentar tanto como regressor como instrumento, apenas para TSLS)
Exemplos:	`add 5 7 9`
	`add xx yy zz --quiet`

Tem que ser invocado após um comando de estimação. As variáveis na *lista-de-variáveis* são acrescentadas ao modelo anterior e o novo modelo é estimado. É apresentada uma estatística de significância conjunta, juntamente com o seu p-value. O teste estatístico é o F no caso de estimação por mínimos quadrados (OLS), ou qui-quadrado assimptótico de Wald nos outros casos. Um p-value abaixo de 0,05 significa que os coeficientes são conjuntamente significantes num nível de 5 porcento.

Se foi fornecida a opção `--quiet` os resultados apresentados ficam confinados ao teste da significância conjunta das variáveis acrescentadas, caso contrário, também serão mostradas as estimativas para o modelo aumentado. Neste caso, a opção `--vcv` faz com que a matriz de covariâncias dos coeficientes também seja apresentada. Se for usada a opção `--silent`, não será mostrado nada; em em todo o caso, os resultados do teste podem ser obtidos usando as variáveis especiais `$test` e `$pvalue`.

Se o modelo original foi estimado usando o método dos mínimos quadrados de duas fases, ocorre uma situação ambígua: devem as novas variáveis serem acrescentadas como sendo regressores, como instrumentos, ou como ambos? Isto resolve-se do seguinte modo: por omissão as novas variáveis são acrescentadas como regressores endógenos, mas se foi dada a opção `--inst` elas são acrescentadas como instrumentos, ou se a opção `--both` está presente elas são acrescentadas como regressores exógenos.

Caminho de Menu: Janela do modelo, /Testes/Acrescentar variáveis

adf

Argumentos:	*ordem nome-de-variável*
Opções:	`--nc` (teste sem constante)
	`--c` (apenas com constante)
	`--ct` (com constante e tendência)
	`--ctt` (com constante, tendência e quadrado da tendência)
	`--seasonals` (incluir variáveis sazonais auxiliares)
	`--verbose` (mostrar resultados da regressão)
	`--quiet` (não mostrar resultados da regressão)
	`--difference` (usar a primeira diferença da variável)
	`--test-down` (ordem 'lag' automática)
Exemplos:	`adf 0 y`
	`adf 2 y --nc --c --ct`
	`adf 12 y --c --test-down`
	Ver também`jgm-1996.inp`

Determina estatísticas para um conjunto de testes de Dickey–Fuller sobre a variável especificada, com a hipótese nula de que a variável tem uma raiz unitária. (Mas se a opção de diferenciação tiver sido dada, a primeira diferença da variável é obtida, e a discussão abaixo deve ser interpretada como sendo referente à variável transformada.)

Por omissão, são apresentadas três variantes do teste: uma baseada na regressão contendo uma constante, uma usando uma constante e uma tendência linear, e uma usando uma constante e uma tendência quadrática. Você pode controlar as variantes que são apresentadas ao especificar uma ou mais opções.

Em todos os casos a variável dependente é a primeira diferença da variável especificada, y, e a variável independente chave é o primeiro 'lag' de y. O modelo é construido de modo a que o coeficiente do 'lag' de y iguale 1 menos a raiz em questão. Por exemplo, o modelo com uma constante pode ser escrito como

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

Se a ordem de 'lag', k, é maior que 0, então k 'lags' da variável dependente são incluidos no lado direito das regressões de teste, sujeitos à seguinte qualificação. Se a opção `--test-down` foi dada, k é considerada como sendo o 'lag' máximo e a ordem de 'lag' efectivamente usada é obtida testando para baixo, de acordo com o seguinte algoritmo:

1. Estimar a regressão de Dickey–Fuller com k 'lags' da variável dependente.
2. O último 'lag' é significante? Se sim, executar o teste com com a ordem de 'lag', k. Senão, fazer k = k − 1; se k for igual a 0, executar o teste com a ordem de 'lag' 0, senão saltar para o passo 1.

No contexto do passo 2 acima, "significante" quer dizer que para o último 'lag', a estatística-t, que segue uma distribuição normal, tem um p-value bilateral assimptótico menor ou igual a 0,10.

Os p-values para os testes de Dickey–Fuller baseiam-se em MacKinnon (1996). O código relevante é incluído com a generosa permissão do autor.

Caminho de Menu: /Variável/Teste de Dickey-Fuller aumentado

append

Argumento:	*ficheiro-de-dados*

Abre um ficheiro de dados e acrescenta esse conteúdo ao conjunto de dados actual, se os novos dados forem compatíveis. O programa tentará determinar o formato do ficheiro de dados (nativo, texto simples, CVS, Gnumeric, Excel, etc.).

Caminho de Menu: /Ficheiro/Acrescentar dados

ar

Argumentos:	*'lags'* ; *variável-dependente variáveis-independentes*
Opção:	`--vcv` (mostrar matriz de covariância)
Exemplo:	`ar 1 3 4 ; y 0 x1 x2 x3`

Determina estimativas para os parâmetros usando o procedimento iteractivo e generalizado de Cochrane–Orcutt (ver a Secção 9.5 de Ramanathan, 2002). A iteração termina quando os erros das somas de quadrados sucessivos não difiram em mais que 0,005 porcento ou após 20 iterações.

'lags' é uma lista de 'lags' nos resíduos, terminada por um ponto-e-vírgula. No exemplo acima o termo do erro é especificado como

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

Caminho de Menu: /Modelo/Série temporal/Estimação autoregressiva

arbond

Argumento:	*p* `[` *q* `]` `;` *variável-dependente variáveis-independentes* `[ ;` *instrumentos* `]`
Opções:	`--vcv` (mostrar matriz de covariância)
	`--two-step` (executa estimação pelo Método Generalizado dos Momentos (GMM) de 2-fases)
	`--time-dummies` (acrescenta variáveis auxiliares tempo)
	`--asymptotic` (erros padrão assimptóticos)
Exemplos:	`arbond 2 ; y Dx1 Dx2`
	`arbond 2 5 ; y Dx1 Dx2 ; Dx1`
	`arbond 1 ; y Dx1 Dx2 ; Dx1 GMM(x2,2,3)`
	Ver também`arbond91.inp`

Executa a estimação de modelos de painel dinâmico (ou seja, modelos de painel que contenham um ou mais 'lags' da variável dependente) recorrendo ao método Método Generalizado dos Momentos (GMM) desenvolvido por Arellano e Bond (1991).

O parâmetro p representa a ordem da autoregressão para a variável dependente. O parâmetro opcional q indica o máximo 'lag' do nível da variável dependente a ser usada como um instrumento. Se este argumento for omitido, ou de valor 0, todos os 'lags' disponíveis são usados.

A variável dependente deve ser dada na forma de níveis; ela será automaticamente diferenciada (pois este estimador usa diferenciação para anular os efeitos individuais). As variáveis independentes não são automaticamente diferenciadas; se você pretende usar diferenças (o que acontece em geral para variáveis quantitativas, mas não será, por exemplo, para variáveis auxiliares temporais), deve primeiro criar essas diferenças e depois especificar estas como sendo regressoras.

O último campo (opcional) do comando serve para especificar instrumentos. Se não for dado nenhum, então é assumido que todas as variáveis independentes são estritamente exógenas. Se você especificar alguns instrumentos, então deve incluir na lista quaisquer variáveis independentes estritamente exógenas. Para regressores predeterminados, você pode usar a função `GMM` para incluir uma gama de 'lags' especificada no modo bloco-diagonal. Isto é ilustrado no terceiro exemplo acima. O primeiro argumento de `GMM` é o nome da variável em questão, o segundo é o mínimo 'lag' a ser usado como instrumento, e o terceiro é o máximo 'lag'. Se o terceiro argumento for dado como 0, todos os 'lags' disponíveis são usados.

Por omissão são apresentados os resultados da estimação 1-fase (com erros padrão robustos). Opcionalmente, você pode escolher estimação de 2-fases. Em ambos os casos são efectuados testes de autocorrelação de ordem 1 e 2, assim como o teste de sobre-identificação de Sargan e o teste de Wald para a significância conjunta dos regressores. Note-se que este modelo de diferenciação com autocorrelação de primeira-ordem não invalida o modelo, mas que a autocorrelação de segunda-ordem não respeita as assunções estatísticas presentes.

No caso da estimação de 2-fases, por omissão, os erros padrão são determinados usando a correcção de amostra-finita sugerida por Windmeijer (2005). Os erros padrão assimptóticos associados ao estimador de 2-fases são em geral considerados como guias para inferência pouco fiáveis, mas se por alguma razão os pretender ver, você pode usar a opção `--asymptotic` para desligar a correcção de Windmeijer.

Se for dada a opção `--time-dummies`, são acrescentadas variáveis auxiliares temporais aos regressores especificados. Para evitar colinearidade perfeita com a constante, o número de auxiliares é uma unidade a menos que o número máximo de períodos usados na estimação. As auxiliares são introduzidas por níveis; se você deseja usar auxiliares de tempo na forma de primeiras-diferenças, você terá que definir e acrescentar essas variáveis manualmente.

Caminho de Menu: /Modelo/Painel/Arellano-Bond

arch

Argumentos:	*ordem variável-dependente variáveis-independentes*
Exemplo:	`arch 4 y 0 x1 x2 x3`

Testa o modelo em ARCH (Heterosquedicidade Condicional Autoregressiva) da ordem de 'lag' especificada. Se a estatística de teste LM tiver um p-value abaixo de 0,10, então a estimação ARCH também é executada. Se a variância predita de qualquer observação na regressão auxiliar não for positiva, então é usado o correspondente resíduo ao quadrado. Segue-se uma estimação por Mínimos Quadrados com Pesos sobre o modelo original.

Ver também garch.

Caminho de Menu: Janela do modelo, /Testes/ARCH

arima

Argumentos:	p d q [; P D Q] ; *variável-dependente* [*variáveis-independentes*]
Opções:	`--verbose` (mostrar detalhes das iterações)
	`--vcv` (mostrar matriz de covariância)
	`--nc` (não incluir uma constante)
	`--conditional` (usar verosimilhança máxima condicional)
	`--x-12-arima` (usar X-12-ARIMA para estimação)
Exemplos:	`arima 1 0 2 ; y`
	`arima 2 0 2 ; y 0 x1 x2 --verbose`
	`arima 0 1 1 ; 0 1 1 ; y --nc`

Se não for dada a lista de *variáveis-independentes*, é estimado um modelo ARIMA (Média Móvel, Autoregressiva, Integrada) univariado. O valores inteiros p, d e q representam respectivamente, a ordem autoregressiva (AR), a ordem de diferenciação, e ordem da média móvel (MA). Estes valores podem ser fornecidos na forma numérica, ou como nome de variáveis escalares pré-existentes. Por exemplo, um valor de 1 em d, significa que a primeira diferença da variável dependente deve ser obtida antes de estimar os parâmetros ARMA.

Os valores inteiros opcionais,P, D e Q representam respectivamente, a sazonalidade AR, a ordem para diferenciação de sazonalidade e a ordem de sazonalidade MA. Estes são apenas aplicáveis se os dados tiverem uma frequência superior a 1 (por exemplo, quadrimestral ou mensal). Mais uma vez, estas ordens podem ser dadas na forma numérica ou como variáveis.

No caso univariado é incluído no modelo por omissão, um interceptor, mas isto pode ser suprimido com a opção `--nc`. Se forem fornecidas *variáveis-independentes*, o modelo passa a ser ARMAX; neste caso a constante deve ser explicitamente incluída se você pretender um interceptor (tal como no segundo exemplo acima).

Existe outra forma alternativa para este comando: se você não pretende aplicar diferenciação (seja sazonal ou não-sazonal), você pode omitir ambos os parâmetros d e D, em vez de entrar explicitamente zeros. Além disso, `arma` é um sinónimo ou aliás para `arima`. Assim, por exemplo, o comando seguinte é válido para especificar o modelo ARMA(2, 1):

```
arma 2 1 ; y
```

O normal é usar a funcionalidade “nativa” gretl ARMA, com estimação de Máxima Verosimilhança (ML) exacta (usando o filtro de Kalman). Outras opções são: código nativo, ML condicional; X-12-ARIMA, ML exacta; e X-12-ARIMA, ML condicional. (As últimas duas opções estão disponíveis apenas se o programa X-12-ARIMA estiver instalado.) Para detalhes sobre estas opções, veja por favor *Manual de Utilização do Gretl*.

O valor AIC retornado em ligação com os modelos ARIMA é calculado conforme a definição usada no programa X-12-ARIMA, nomeadamente

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

onde ℓ é o logaritmo da verosimilhança e k é o número total de parâmetros estimados. Note-se que o programa X-12-ARIMA não produz critérios de informação tal como o AIC quando a estimação é por ML condicional.

A imagem da “frequência” apresentada em ligação com as raízes AR e MA é valor λ que resolve

$$z = re^{i2\pi\lambda}$$

onde z é a raiz em questão e r o seu módulo.

Caminho de Menu: /Modelo/Série temporal/ARIMA

Acesso alternativo: Menu de contexto da janela principal (selecção singular)

boxplot

Argumento:	*lista-de-variáveis*
Opção:	`--notches` (mostrar intervalo de 90 porcento para a mediana delimitado por entalhes)

Estes gráficos (criados por Tukey e Chambers) apresentam a distribuição de uma variável. A caixa central contém os 50 porcento dos dados centrais, i.e. está limitada pelos primeiro e terceiro quartis. Os "bigodes" estendem-se até aos valores mínimo e máximo. É desenhada uma linha que corta a caixa na mediana.

No caso de caixas com entalhes, os entalhes representam os limites do intervalo de confiança para a mediana de cerca de 90 porcento. Isto é obtido usando o método 'bootstrap'.

A seguir a cada variável indicada no comando caixa-com-bigodes, pode-se acrescentar uma expressão Booleana para restringir a variável em questão. Tem que se inserir um espaço entre o nome da variável ou número, e a expressão. Suponha que você dispõe de valores de salários (`salary`) para homens e mulheres, e que tem a variável auxiliar `GENDER` com valor 1 para homens e 0 para mulheres. Nesse caso você podia ter gráficos caixa-com-bigodes comparativos com a seguinte *lista-de-variáveis*:

```
salary (GENDER=1) salary (GENDER=0)
```

Alguns detalhes das caixas-com-bigodes do gretl podem ser controlados por intermédio de um ficheiro (de texto simples) com o nome `.boxplotrc`. Para mais detalhes sobre isto veja *Manual de Utilização do Gretl*.

Caminho de Menu: /Ver/Gráfico das variáveis/Caixa com bigodes

break

Sai de um ciclo. Este comando pode apenas ser usado dentro de um ciclo; ele termina a execução de comandos e sai de dentro do ciclo (o mais interior). Ver também loop.

chow

Argumento:	*obs*
Exemplos:	`chow 25`
	`chow 1988:1`

Tem que se seguir a uma regressão de Mínimos Quadrados (OLS). Cria uma variável auxiliar que é igual a 1 a partir do ponto especificado por *obs* até ao final da amostra, caso contrário é 0, e cria também termos de interacção entre esta variável auxiliar e as variáveis independentes originais. É executada uma regressão aumentada que inclui estes termos e é calculada a estatística F, considerando a regressão aumentada como não restringida e a original como restringida. Esta estatística é apropriada para testar a hipótese nula de não existência de quebra estrutural no ponto de separação dado.

Caminho de Menu: Janela do modelo, /Testes/Teste de Chow

coeffsum

Argumento:	*lista-de-variáveis*
Exemplo:	`coeffsum xt xt_1 xr_2`
	`restrict.inp`

Tem que se seguir a uma regressão. Calcula a soma dos coeficientes nas variáveis indicadas na *lista-de-variáveis*. Apresenta esta soma juntamente com o seu erro padrão e o p-value para a hipótese nula de que a soma é zero.

Note-se a diferença entre este teste e omit, que testa a hipótese nula de que os coeficientes num conjunto especificado de variáveis independentes são *todos* iguais a zero.

Caminho de Menu: Janela do modelo, /Testes/Soma de coeficientes

coint

Argumentos:	*ordem variável-dependente variáveis-independentes*
Opções:	`--nc` (não incluir uma constante)
	`--ct` (incluir constante e tendência)
	`--ctt` (incluir constante e tendência quadrática)
	`--skip-df` (não efectuar testes DF nas variáveis individuais)
Exemplos:	`coint 4 y x1 x2`
	`coint 0 y x1 x2 --ct --skip-df`

O teste de cointegração Engle–Granger. O procedimento por omissão é: (1) efectuar testes de Dickey–Fuller (DF) segundo a hipótese nula de que cada variável listada tem uma raiz unitária; (2) estima a regressão de cointegração; e (3) executar um teste DF sobre os resíduos da regressão de cointegração. Se for dada a opção `--skip-df`, o passo (1) é omitido.

Se a ordem de 'lag' especificada é positiva, todos os testes Dickey–Fuller usam essa ordem. Se a ordem for antecedida de um sinal menos, ela é encarada como sendo o máximo 'lag' e a ordem efectivamente usada em cada caso é obtida testando para baixo: ver o comando adf para detalhes.

Por omissão, a regressão de cointegração contém uma constante. Se você deseja suprimir a constante, acrescente a opção `--nc`. Se você deseja aumentar a lista de termos determinísticos na regressão de cointegração com uma tendência linear ou quadrática, use a opção `--ct` ou `--ctt`. Estas opções são mutualmente exclusivas.

Os *P*-values para este teste são baseados em MacKinnon (1996). O código relevante é incluído com a generosa permissão do autor.

Caminho de Menu: /Modelo/Série temporal/Testes de Cointegração/Engle-Granger

coint2

Argumentos:	*ordem variável-dependente variáveis-independentes*
Opções:	`--nc` (sem constante)
	`--rc` (constante restringida)
	`--crt` (constante e tendência restringida)
	`--ct` (constante e tendência não restringida)
	`--seasonals` (incluir auxiliares sazonais centradas)
	`--quiet` (apenas mostrar os testes)
	`--verbose` (mostrar detalhes das regressões auxiliares)
Exemplos:	`coint2 2 y x`
	`coint2 4 y x1 x2 --verbose`
	`coint2 3 y x1 x2 --rc`

Executa o teste de Johansen para a cointegração entre as variáveis listadas para a dada ordem de 'lag'. Os valores críticos são determinados usando a aproximação gamma de J. Doornik (Doornik, 1998). Para detalhes sobre este teste ver o Capítulo 20 do livro de Hamilton, *Time Series Analysis* (1994).

A inclusão de termos determinísticos no modelo é controlada por intermédio das opções. Por omissão, se não tiver sido indicada nenhuma opção, será incluída uma "constante não restringida", o que permite a presença de um interceptor não-nulo nas relações cointegrantes assim como uma tendência nos níveis das variáveis endógenas. Na literatura derivada do trabalho de Johansen (ver por exemplo o livro dele de 1995) isto é frequentemente referido como sendo o "caso 3". As primeiras quatro opções apresentadas acima, que são mutualmente exclusivas, produzem respectivamente os casos 1, 2, 4, e 5. O significado destes casos e os critérios para seleccionar um caso estão explicados no *Manual de Utilização do Gretl*.

A opção `--seasonals`, que pode ser combinada com qualquer outra opção, especifica a inclusão de um conjunto de variáveis auxiliares sazonais. Esta opção apenas está disponível para dados trimestrais ou mensais.

A seguinte tabela serve como uma guia à interpretação dos resultados apresentados pelo teste, num caso de 3-variáveis. `H0` significa a hipótese nula, `H1` a hipótese alternativa, e `c` o número de relações cointegrantes.

```
Ordem    Teste Traço        Teste Lmax
         H0     H1          H0     H1
---------------------------------------
 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
---------------------------------------
```

Ver também o comando vecm.

Caminho de Menu: /Modelo/Série temporal/Testes de Cointegração/Johansen

corc

Argumentos:	*variável-dependente variáveis-independentes*
Opção:	`--vcv` (mostrar matriz de covariância)
Exemplo:	`corc 1 0 2 4 6 7`

Calcula estimativas dos parâmetros usando o procedimento iterado de Cochrane–Orcutt (ver na Secção 9.4 de Ramanathan, 2002). As iterações terminam quando estimativas sucessivas do coeficiente de autocorrelação diferem menos que 0,001 ou após 20 iterações.

Caminho de Menu: /Modelo/Série temporal/Cochrane-Orcutt

corr

Argumento:	`[` *lista-de-variáveis* `]`
Exemplo:	`corr y x1 x2 x3`

Apresenta os coeficientes de correlação emparelhados das variáveis em *lista-de-variáveis*, ou de todas as variáveis no conjunto de dados se não for dada a *lista-de-variáveis*.

Caminho de Menu: /Ver/Matriz de correlação

Acesso alternativo: Menu de contexto da janela principal (selecção múltipla)

corrgm

Argumentos:	*variável* `[` *maxlag* `]`
Exemplo:	`corrgm x 12`

Apresenta os valores da função de autocorrelação para a *variável*, que pode ser especificada por nome ou por número. Os valores são definidos pela equação $\hat{\rho}(u_t, u_{t-s})$ onde u_t é a t–ésima observação da variável u e s é o número de "lags".

Também são apresentadas as autocorrelações parciais (obtidas segundo o algoritmo de Durbin–Levinson): estas constituem a rede dos efeitos dos "lags"intervenientes. O comando também produz o gráfico correlograma e apresenta a estatística de teste Q de Box–Pierce, para a hipótese nula de que a série temporal é "ruído branco": terá uma distribuição qui-quadrado assimptótico com os graus de liberdade iguais ao número de "lags"usados.

Se o valor *maxlag* for especificado o comprimento do correlograma fica limitado a esse máximo número de "lags", senão o comprimento é determinado automáticamente, como uma função da frequência dos dados e do número de observações.

Caminho de Menu: /Variável/Correlograma

Acesso alternativo: Menu de contexto da janela principal (selecção singular)

criteria

Argumentos:	*ess T k*
Exemplo:	`criteria 23.45 45 8`

Determina o Critério de Informação de Akaike (AIC) e o Critério de Informação Bayesiano de Schwarz (BIC), dados *ess* (erro da soma de quadrados), o número de observações (T), e o número de coeficientes (k). T, k, e *ess* podem ser valores numéricos ou nomes de variáveis préviamente definidas.

O AIC é obtido segundo a formulação original de Akaike (1974), nomeadamente

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

onde ℓ designa a verosimilhança logaritmica maximizada. O BIC é calculado por

$$\mathrm{BIC} = -2\ell + k \log T$$

Por favor consulte *Manual de Utilização do Gretl* para mais pormenores.

cusum

Opção: `--squares` (executa o teste CUSUMSQ)

Tem que se seguir à estimação de um modelo por via de OLS. Executa o teste CUSUM —ou se for dada a opção `--squares`, o teste CUSUMSQ —para a estabilidade dos parâmetros. É obtida uma série temporal de erros de predição um passo-à-frente, pela execução de séries de regressões: a primeira regressão usa as primeiras k observações e é usada para gerar a predição da variável dependente na observação k + 1; a segunda usa a primeira predição para a observação k + 2, e por aí a diante (onde k é o número de parâmetros no modelo original).

A soma acumulada dos erros de predição escalados, ou os quadrados desses erros, é mostrada e apresentada em gráfico. A hipótese nula para a estabilidade dos parâmetros é rejeitada ao nível de cinco porcento, se a soma acumulada se desviar do intervalo de confiança de 95 porcento.

No caso do teste CUSUM, é também apresentada a estatística de teste t de Harvey–Collier, para a hipótese nula da estabilidade dos parâmetros. Ver o livro *Econometric Analysis* de Greene para mais detalhes. Para o teste CUSUMSQ, o intervalo de confiança a 95 porcento é calculado de acordo com o algoritmo apresentado por Edgerton e Wells (1994).

Caminho de Menu: Janela do modelo, /Testes/Teste CUSUM(SQ)

data

Argumento: *lista-de-variáveis*

Reads the variables in *lista-de-variáveis* 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.

Caminho de Menu: /Ficheiro/Databases

dataset

Argumentos: *keyword parameters*

Exemplos:
```
dataset addobs 24
dataset compact 1
dataset compact 4 last
dataset expand 12
dataset transpose
dataset sortby x1
```

Performs various operations on the data set as a whole, depending on the given *keyword*, which must be `addobs`, `compact`, `expand`, `transpose`, `sortby` or `dsortby`. Note: 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.

`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`: Must be followed by a positive integer representing a new data frequency, which should be higher than the current frequency. This command is only available for annual or quarterly time series data. Annual data can be expanded to quarterly or monthly; quarterly data can be expanded to monthly. All the series in the data set are padded out to the new frequency by repeating the existing values.

`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`: One variable name is required; this variable is used as a sort key. The observations on all variables in the dataset are re-ordered by increasing value of the key variable. This command is available only for undated data.

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

Caminho de Menu: /Dados/Acrescentar observações

Caminho de Menu: /Dados/Compactar datos

Caminho de Menu: /Dados/Expandir datos

Caminho de Menu: /Dados/Transpôr

Caminho de Menu: /Dados/Ordenar

delete

Argumento: [*lista-de-variáveis*]

Removes the listed variables (given by name or number) from the dataset. *Use with caution*: no confirmation is asked, and any variables with higher ID numbers will be re-numbered.

If no *lista-de-variáveis* is given with this command, it deletes the last (highest numbered) variable from the dataset.

Caminho de Menu: Main window pop-up (single selection)

diff

Argumento: *lista-de-variáveis*

The first difference of each variable in *lista-de-variáveis* 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)
```

Caminho de Menu: /Add/First differences of selected variables

difftest

Argumentos: *var1 var2*
Opções: `--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

Argumento: *lista-de-variáveis*
Opção: `--reverse` (mark variables as continuous)

Marks each variable in *lista-de-variáveis* 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 *lista-de-variáveis* are marked as being continuous.

Caminho de Menu: /Variável/Edit attributes

dummify

Argumento: *lista-de-variáveis*
Opções: `--drop-first` (omit lowest value from encoding)
`--drop-last` (omit highest value from encoding)

For any suitable variables in *lista-de-variáveis*, 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)
```

else

See if.

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

Argumento: [*-f filename*]
Opção: `--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.

Caminho de Menu: Janela do modelo, /LaTeX

equation

Argumentos: *variável-dependente variáveis-independentes*
Exemplo: `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

Argumentos:	*systemname estimator*
Opções:	`--iterate` (iterate to convergence)
	`--no-df-corr` (no degrees of freedom correction)
	`--geomean` (see below)
Exemplos:	`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=`.

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 ks denote the number of independent parameters in each equation.

fcast

Argumentos:	`[` *startobs endobs* `]` *fitvar*
Opções:	`--dynamic` (create dynamic forecast)
	`--static` (create static forecast)
Exemplos:	`fcast 1997:1 2001:4 f1`
	`fcast fit2`

Must follow an estimation command. Forecasts are generated for the specified range (or the largest possible range if no *startobs* and *endobs* are given) and the values saved as *fitvar*, which can be printed, graphed, or plotted. The right-hand side variables are those in the original model. There is no provision to substitute other variables. If an autoregressive error process is specified the forecast incorporates the predictable fraction of the error process.

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. See fcasterr for more details.

Caminho de Menu: Janela do modelo, /Analysis/Forecasts

fcasterr

Argumentos:	*startobs endobs*
Opções:	`--plot` (display graph)
	`--dynamic` (create dynamic forecast)
	`--static` (create static forecast)

After estimating a model you can use this command to print out fitted values over the specified observation range, along with (depending on the nature of the model and the available data) estimated standard errors of those predictions and 95 percent confidence intervals.

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 with 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 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.

Caminho de Menu: Janela do modelo, /Analysis/Forecasts

fit

A shortcut to fcast. Must follow an estimation command. Generates fitted values, in a series called `autofit`, for the current sample, based on the last regression. In the case of time-series models, also pops up a graph of fitted and actual values of the dependent variable against time.

freq

Argumento:	*var*
Opções:	`--quiet` (suppress printing of histogram)
	`--gamma` (test for gamma distribution)

With no options given, displays the frequency distribution for *var* (given by name or number) and shows the results of the Doornik–Hansen chi-square test for normality.

If the `--quiet` option is given, the histogram is not shown. 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).

In interactive mode a graph of the distribution is displayed.

Caminho de Menu: /Variável/Frequency distribution

function

Argumento:	*fnname*

Opens a block of statements in which a function is defined. This block must be closed with `end function`. Please see *Manual de Utilização do Gretl* for details.

garch

Argumentos:	*p q* ; *variável-dependente* [*variáveis-independentes*]
Opções:	`--robust` (robust standard errors)
	`--verbose` (mostrar detalhes das iterações)
	`--vcv` (mostrar matriz de covariância)
	`--arma-init` (initial variance parameters from ARMA)
Exemplos:	`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 *variáveis-independentes* 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:

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

The gretl GARCH algorithm is basically that of Fiorentini, Calzolari and Panattoni (1996), used by kind permission of Professor Fiorentini.

Several variant estimates of the coefficient 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:

```
genr ht = $h
```

Caminho de Menu: /Modelo/Série temporal/GARCH

genr

Argumentos:	*newvar* = *formula*

Creates new variables, usually through transformations of existing variables. See also diff, logs, lags, ldiff, multiply, 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 *Manual de Utilização do Gretl*.

This 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 or vector 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 $\texttt{x} > 10$, 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`.

Supported *functions* fall into these groups:

- Standard mathematical functions: abs, exp, int (integer part), ln (natural logarithm: log is a synonym), log10 (log to the base 10), log2 (log to the base 2), sqrt. All of these take a single argument, which may be either a series or a scalar.
- Trigonometric functions: cos, sin, tan and atan (arc tangent). These functions also take a single series or scalar argument.
- Standard statistical functions taking a single argument and yielding a scalar result: max (maximum value in a series), min (minimum value in series), mean (arithmetic mean), median, var (variance), sd (standard deviation), sst (sum of squared deviations from the mean), sum, gini (Gini coefficient).
- Statistical functions taking one series as argument and yielding a series or vector result: sort (sort a series in ascending order of magnitude), dsort (sort in descending order), cum (cumulate, or running sum). For panel data only, the functions pmean and psd take a single series as argument and return series containing, respectively, the means and standard deviations of the variable for each of the cross-sectional units. See *Manual de Utilização do Gretl* for details.
- Statistical functions taking two series as arguments and yielding a scalar result: cov (covariance), corr (correlation coefficient).
- Special statistical functions: pvalue, cdf and critical (see below), cnorm (standard normal CDF), dnorm (standard normal PDF), qnorm (standard normal quantiles, or inverse of standard normal PDF), resample (resample a series with replacement, for bootstrap purposes), hpfilt (Hodrick–Prescott filter: this function returns the “cycle” component of the series), bkfilt (Baxter–King bandpass filter).
- Time-series functions: diff (first difference), ldiff (log-difference, or first difference of natural logs), sdiff (seasonal difference), fracdiff (fractional difference). To generate lags of a variable `x`, use the syntax x(-N), where `N` represents the desired lag length; to generate leads, use x(+N).
- Dataset functions yielding a series: misszero (replaces the missing observation code in a given series with zeros); zeromiss (the inverse operation to misszero); missing (at each observation, 1 if the argument has a missing value, 0 otherwise); ok (the opposite of missing).
- Dataset functions yielding a scalar: nobs (gives the number of valid observations in a data series), firstobs (gives the 1-based observation number of the first non-missing value in a series), lastobs (observation number of the last non-missing observation in a series).
- Pseudo-random numbers: uniform, normal, chisq, student, binomial and poisson. The first two functions do not take an argument and should be written with empty parentheses: `uniform()`, `normal()`. They create pseudo-random series drawn from the uniform (0–1) and standard normal distributions respectively. The next two take a single argument, namely the (integer) degrees of freedom; they generate drawings from the Chi-square and Student’s t distributions respectively. The binomial function takes two parameters: an integer number of trials and a “success” probability. The poisson function takes a single argument, the mean, which may be either a scalar or a series. Uniform series, which form the basis for the more complex distributions, are generated using the Mersenne Twister;See Matsumoto and Nishimura (1998). The implementation is provided by glib, if available, or by the C code written by Nishimura and Matsumoto.

for normal series the method of Box and Muller (1958) is used, taking input from the Mersenne Twister. Ver também o comando set command, `seed` option.

With a few exceptions as noted, all the above functions take as their single argument either the name of a variable or an expression that evaluates to a variable (e.g. ln((x1+x2)/2)).

The pvalue, critical and cdf functions take, as their first parameter, a one-character code for the distribution of interest:

- z or N: Normal;
- t: Student's t;
- X: chi-square;
- F: Snedecor;
- G: Gamma (only pvalue and cdf);
- B: Binomial (only pvalue and cdf);
- D: Bivariate normal (only cdf).

The pvalue function takes the same arguments as the pvalue command, but in this context commas should be placed between the arguments. It returns a one-tailed, right-hand p-value, that is, the integral of the given density function from the specified value to plus infinity.

The cdf function also takes the same arguments as the pvalue command, but it returns the complementary value, that is, the integral of the relevant density function from its lower limit (either minus infinity or zero) to the specified value. For the bivariate normal distribution, three arguments must be given: x, y and the correlation coefficient *rho*, in this order. For additional discussion of these functions see *Manual de Utilização do Gretl*.

The critical function returns the critical value for a specified probability distribution and a specified proportion in the right-hand tail (with specified degrees of freedom where applicable). The last parameter is the right-hand tail proportion. If the first parameter is t or X, a second parameter must give the degrees of freedom. For the F distribution, the second and third parameters must give the numerator and denominator degrees of freedom.

Here are some examples of use of the pvalue and critical functions (spaces between the arguments are optional):

```
genr p1 = pvalue(z, 2.2)
genr p2 = pvalue(X, 3, 5.67)
genr p3 = cdf(D, -1, 1, 0.25)
genr c1 = critical(t, 20, 0.025)
genr c2 = critical(F, 4, 48, 0.05)
```

The functions relating to the standard normal density work thus, for a given argument x: cnorm returns the area under the standard normal density function integrated from minus infinity to x; dnorm returns the standard normal density evaluated at x; and qnorm returns the z such that the area under the standard normal density, integrated from minus infinity to z, equals x.

The fracdiff function takes two arguments: the name of the series and a fraction, in the range -1 to 1.

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. For example, in the case of quarterly data (periodicity 4), the program creates `dummy_1` = 1 for first quarter and 0 in other quarters, `dummy_2` = 1 for the second quarter and 0 in other quarters, and so on.

- 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.

Various internal variables defined in the course of running a regression can be retrieved using genr, as follows:

- `$ess`: error sum of squares
- `$rsq`: unadjusted R-squared
- `$T`: number of observations used
- `$df`: degrees of freedom
- `$ncoeff`: total number of estimated coefficients
- `$trsq`: TR-squared (sample size times R-squared)
- `$sigma`: standard error of residuals
- `$aic`: Akaike Information Criterion
- `$bic`: Schwarz's Bayesian Information Criterion
- `$hqc`: Hannan–Quinn Criterion
- `$lnl`: log-likelihood (where applicable)
- `$coeff`(*var*): estimated coefficient for variable *var*
- `$stderr`(*var*): estimated standard error for variable *var*
- `$rho`(i): ith order autoregressive coefficient for residuals
- `$vcv`(*x1*,*x2*): estimated covariance between coefficients for the variables *x1* and *x2*

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 internal series `$uhat` and `$yhat` hold, respectively, the residuals and fitted values from the last regression. For GARCH models, the conditional variance is held in the internal variable `$h`.

Three "internal" dataset variables are available: `$nobs` holds the number of observations in the current sample range (which may or may not equal the value of `$T`, the number of observations used in estimating the last model); `$nvars` holds the number of variables in the dataset (including the constant); and `$pd` holds the frequency or periodicity of the data (e.g. 4 for quarterly data).

Two special internal scalars, `$test` and `$pvalue`, hold respectively the value and the p-value of the test statistic that was generated by the last explicit hypothesis-testing command, if any (e.g. `chow`). Please see *Manual de Utilização do Gretl* for details on this.

The variable `t` serves as an index of the observations. For instance `genr dum = (t=15)` will generate a dummy variable that has value 1 for observation 15, 0 otherwise. The variable `obs` is similar but more flexible: you can use this to pick out particular observations by date or name. For example, `genr d = (obs>1986:4)` or `genr d = (obs="CA")`. The last form presumes that the observations are labeled; the label must be put in double quotes.

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`.

Here are a couple of tips on dummy variables:

- Suppose `x` is coded with values 1, 2, or 3 and you want three dummy variables, `d1` = 1 if `x` = 1, 0 otherwise, `d2` = 1 if `x` = 2, and so on. To create these, use the commands:

```
genr d1 = (x=1)
genr d2 = (x=2)
genr d3 = (x=3)
```

- To create `z` = `max(x,y)` do

```
genr d = x>y
genr z = (x*d)+(y*(1-d))
```

Tabela 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 −999 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.

Caminho de Menu: /Variável/Define new variable

Acesso alternativo: Menu de contexto da janela principal

gmm

Opções: `--two-step` (two step estimation)
`--iterate` (iterated GMM)
`--vcv` (print covariance matrix)
`--verbose` (mostrar detalhes das iterações)

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 (tyically, 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`.

Please see *Manual de Utilização do Gretl* 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`.

Caminho de Menu: /Model/GMM

gnuplot

Argumentos:	*yvars xvar* [*dumvar*]
Opções:	`--with-lines` (use lines, not points)
	`--with-impulses` (use vertical lines)
	`--suppress-fit` (don't show fitted line)
	`--linear-fit` (show least squares fit)
	`--inverse-fit` (show inverse fit)
	`--quadratic-fit` (show quadratic fit)
	`--loess-fit` (show loess fit)
	`--dummy` (see below)
Exemplos:	`gnuplot y1 y2 x`
	`gnuplot x time --with-lines`
	`gnuplot wages educ gender --dummy`

Without the `--dummy` option, the *yvars* are graphed against *xvar*. With `--dummy`, *yvar* is graphed against *xvar* with the points shown in different colors depending on whether the value of *dumvar* is 1 or 0.

The `time` variable behaves specially: if it does not already exist then it will be generated automatically.

In interactive mode the result is displayed immediately. In batch mode a gnuplot command file is written, 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).

The various "fit" options are applicable only in the case of a bivariate scatterplot. The default behavior is to show the OLS fitted line if and only if the slope coefficient is significant at the 10 percent level. If the `suppress` option is given, no fitted line is shown. If the `linear` option is given, the OLS line is shown regardless of whether or not it is significant. The other options—`inverse`, `quadratic` and loess—produce respectively an inverse fit (regression of y on 1/x), a quadratic fit, or a loess fit. Loess (also sometimes called "lowess") is a robust locally weighted regression.

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]; }
```

Caminho de Menu: /View/Graph specified vars

Acesso alternativo: Menu de contexto da janela principal, graph button on toolbar

graph

Argumentos:	*yvars xvar*
Opção:	`--tall` (use 40 rows)

ASCII graphics. The *yvars* (which may be given by name or number) are graphed against *xvar* using ASCII symbols. The `--tall` flag will produce a graph with 40 rows and 60 columns. Without it, the graph will be 20 by 60 (for screen output). See also gnuplot.

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.

Caminho de Menu: Janela do modelo, /Testes/Panel diagnostics

hccm

Argumentos:	*variável-dependente variáveis-independentes*
Opção:	`--vcv` (mostrar matriz de covariância)

Heteroskedasticity-Consistent Covariance Matrix: this command runs a regression where the coefficients are estimated via the standard OLS procedure, but the standard errors of the coefficient estimates are computed in a manner that is robust in the face of heteroskedasticity, namely using the MacKinnon–White "jackknife" procedure.

heckit

Argumentos:	*variável-dependente variáveis-independentes* ; *selection equation*
Opções:	`--two-step` (perform two-step estimation)
	`--vcv` (print covariance matrix)
	`--verbose` (print extra output)
Exemplo:	`heckit y 0 x1 x2 ; ys 0 x3 x4`

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).

FIXME this entry needs to be completed.

Caminho de Menu: /Model/Nonlinear models/Heckit

help

Gives a list of available commands. help *command* describes *command* (e.g. help smpl). You can type man instead of help if you like.

Caminho de Menu: /Help

hilu

Argumentos:	*variável-dependente variáveis-independentes*
Opções:	`--vcv` (mostrar matriz de covariância)
	`--no-corc` (do not fine-tune results with Cochrane-Orcutt)

Computes parameter estimates for the specified model using the Hildreth–Lu search procedure. The results are fine-tuned using the Cochrane–Orcutt iterative method, unless the `--no-corc` flag is specified.

This procedure is designed to correct for serial correlation of the error term. The error sum of squares of the transformed model is graphed against the value of rho from −0.99 to 0.99.

Caminho de Menu: /Modelo/Série temporal/Hildreth-Lu

hsk

Argumentos:	*variável-dependente variáveis-independentes*
Opção:	`--vcv` (mostrar matriz de covariância)

An OLS regression is run and the residuals are saved. The logs of the squares of these residuals then become the dependent variable in an auxiliary regression, on the right-hand side of which are the original independent variables plus their squares. The fitted values from the auxiliary regression are then used to construct a weight series, and the original model is re-estimated using weighted least squares. This final result is reported.

The weight series is formed as $1/\sqrt{e^{y^*}}$, where y* denotes the fitted values from the auxiliary regression.

Caminho de Menu: /Model/Other linear models/Heteroskedasticity corrected

hurst

Argumento:	*nome-de-variável*

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.

Caminho de Menu: /Variável/Hurst exponent

if

Flow control for command execution. The syntax is:

`if` *condition*
  commands1
`else`
  commands2
`endif`

condition must be a Boolean expression, for the syntax of which see genr. The else block is optional; `if` ... `endif` blocks may be nested.

include

Argumento:	*inputfile*
Exemplos:	`include myfile.inp`
	`include sols.gfn`

Intended for use in a command script, primarily for including definitions of functions. Executes the commands in *inputfile* 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.

Caminho de Menu: /Data/Read info

Acesso alternativo: Data browser windows

kpss

Argumentos:	*ordem nome-de-variável*
Opções:	`--trend` (include a trend)
	`--verbose` (mostrar resultados da regressão)
	`--quiet` (não mostrar resultados da regressão)
	`--difference` (usar a primeira diferença da variável)
Exemplos:	`kpss 8 y`
	`kpss 4 x1 --trend`

Computes the KPSS test (Kwiatkowski, Phillips, Schmidt and Shin, 1992) for stationarity of the specified variable (or its 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.

Caminho de Menu: /Variável/KPSS test

labels

Prints out the informative labels for any variables that have been generated using genr, and any labels added to the data set via the GUI.

lad

Argumentos:	*variável-dependente variáveis-independentes*
Opção:	`--vcv` (mostrar matriz de covariância)

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.

Caminho de Menu: /Model/Robust estimation/Least Absolute Deviation

lags

Variantes:	`lags` *lista-de-variáveis*
	`lags` *order* `;` *lista-de-variáveis*
Exemplos:	`lags x y`
	`lags 12 ; x y`

Creates new variables which are lagged values of each of the variables in *lista-de-variáveis*. By default the number of lagged variables 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.

Caminho de Menu: /Add/Lags of selected variables

ldiff

Argumento: *lista-de-variáveis*

The first difference of the natural log of each variable in *lista-de-variáveis* is obtained and the result stored in a new variable 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))
```

Caminho de Menu: /Add/Log differences of selected variables

leverage

Opção: `--save` (save variables)

Must immediately 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 shown: these are "studentized residuals" (predicted residuals divided by their standard errors) multiplied by $\sqrt{h/(1-h)}$. For a discussion of studentized residuals and DFFITS see G. S. Maddala, *Introduction to Econometrics*, chapter 12; also Belsley, Kuh and Welsch (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.

Caminho de Menu: Janela do modelo, /Testes/Influential observations

lmtest

Argumento: [*order*]
Opções: `--logs` (non-linearity, logs)
`--autocorr` (serial correlation)
`--arch` (ARCH)
`--squares` (non-linearity, squares)
`--white` (heteroskedasticity, White's test)
`--panel` (heteroskedasticity, groupwise)
`--quiet` (don't print auxiliary regression)

Must immediately follow an ols command. Depending on the options given, this command carries out some combination of the following: Lagrange Multiplier tests for nonlinearity (logs and squares); White's test for heteroskedasticity; the LMF test for serial correlation (see Kiviet, 1986); and a test for ARCH (Autoregressive Conditional Heteroskedasticity, see also the arch command).

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 `--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).

By default, the program prints the auxiliary regression on which the test statistic is based. This may be suppressed by using the `--quiet` flag.

Caminho de Menu: Janela do modelo, /Tests

logistic

Argumentos: *variável-dependente variáveis-independentes* [`ymax=`*value*]
Opção: `--vcv` (mostrar matriz de covariância)
Exemplos: `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 it is a decimal fraction, between 0 and 1, the default is to use a y* value (the asymptotic maximum of the dependent variable) of 1. If the dependent variable is a percentage, between 0 and 100, the default y* is 100.

If you wish to set a different maximum, use the optional `ymax=`*value* syntax following the list of regressors. 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.

Caminho de Menu: /Model/Nonlinear models/Logistic

logit

Argumentos:	*variável-dependente variáveis-independentes*
Opções:	`--robust` (robust standard errors)
	`--vcv` (mostrar matriz de covariância)
	`--verbose` (mostrar detalhes das iterações)

If the dependent variable is a binary variable (all values are 0 or 1) maximum likelihood estimates of the coefficients on *variáveis-independentes* are obtained via the "binary response model regression" (BRMR) method outlined by Davidson and MacKinnon (2004). As the model is nonlinear the slopes depend on the values of the independent variables: the reported slopes are evaluated at the means of those variables. 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 Davidson and MacKinnon (2004, Chapter 10) for details.

If the dependent variable is not binary, but is discrete and has a minimum value of 0, then Ordered Logit estimates are obtained. In this case robust standard errors are not yet available. (If the variable selected as dependent is not discrete, or does not have a minimum of zero, an error is flagged.)

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

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

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

Caminho de Menu: /Model/Nonlinear models/Logit

logs

Argumento:	*lista-de-variáveis*

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

Caminho de Menu: /Add/Logs of selected variables

loop

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

The parameter *control* must take one of four forms, as shown in the examples: an integer number of times to repeat the commands within the loop; "`while`" plus a numerical condition; "`for`" plus a range of values for the internal index variable `i`; or "`for`" plus three expressions in parentheses, separated by semicolons. In the last form the left-hand expression initializes a variable, the middle expression sets a condition for iteration to continue, and the right-hand expression sets an increment or decrement to be applied at the start of the second and subsequent iterations. (This is a restricted form of the `for` statement in the C programming language.)

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.

See *Manual de Utilização do Gretl* 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

Argumento:	*lista-de-variáveis*
Opções:	`--save` (add distances to the dataset)
	`--vcv` (mostrar matriz de covariância)

The Mahalanobis distance is the distance between two points in an 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).

Caminho de Menu: /View/Mahalanobis distances

meantest

Argumentos:	*var1 var2*
Opção:	`--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.

Caminho de Menu: /Model/Bivariate tests/Difference of means

mle

Argumentos:	*log-likelihood function derivatives*
Opções:	`--vcv` (mostrar matriz de covariância)
	`--hessian` (base covariance matrix on the Hessian)
	`--robust` (QML covariance matrix)
	`--verbose` (mostrar detalhes das iterações)
Exemplo:	`weibull.inp`

Performs Maximum Likelihood (ML) estimation using the BFGS (Broyden, Fletcher, Goldfarb, Shanno) algorithm. 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. This can be done using the genr command. For example, `genr 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. 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).

Caminho de Menu: /Model/Maximum likelihood

modeltab

Argumentos:	*add* ou *show* ou *free*

Manipulates the gretl "model table". See *Manual de Utilização do Gretl* for details. The subcommands 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.

Caminho de Menu: Session window, Model table icon

mpols

Argumentos:	*variável-dependente variáveis-independentes*
Opções:	`--vcv` (mostrar matriz de covariância)
	`--simple-print` (do not print auxiliary statistics)
	`--quiet` (não mostrar resultados da regressão)

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 *variáveis-independentes* 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 *variáveis-independentes*. 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
```

Caminho de Menu: /Model/Other linear models/High precision OLS

multiply

Argumentos:	*x suffix lista-de-variáveis*
Exemplos:	`multiply invpop pc 3 4 5 6`
	`multiply 1000 big x1 x2 x3`

The variables in *lista-de-variáveis* (referenced by name or number) are multiplied by *x*, which may be either a numerical value or the name of a variable already defined. The products are named with the specified *suffix* (maximum 3 characters). The original variable names are truncated first if need be. For instance, suppose you want to create per capita versions of certain variables, and you have the variable `pop` (population). A suitable set of commands is then:

```
genr invpop = 1/pop
multiply invpop pc income
```

which will create `incomepc` as the product of `income` and `invpop`, and `expendpc` as `expend` times `invpop`.

nls

Argumentos:	*function* [*derivatives*]
Opções:	`--robust` (robust standard errors)
	`--vcv` (mostrar matriz de covariância)
	`--verbose` (mostrar detalhes das iterações)
Exemplo:	`wg_nls.inp`

Performs Nonlinear Least Squares (NLS) estimation using a modified version of the Levenberg–Marquandt 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   genr a = $coeff(0)
4   genr b = $coeff(Y)
5   genr 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.

For further details on NLS estimation please see *Manual de Utilização do Gretl*.

Caminho de Menu: /Model/Nonlinear models/Nonlinear Least Squares

nulldata

Argumento:	*series-length*
Exemplo:	`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.

Caminho de Menu: /Ficheiro/New data set

ols

Argumentos:	*variável-dependente variáveis-independentes*
Opções:	`--vcv` (mostrar matriz de covariância)
	`--robust` (robust standard errors)
	`--simple-print` (do not print auxiliary statistics)
	`--quiet` (não mostrar resultados da regressão)
	`--no-df-corr` (suppress degrees of freedom correction)
	`--print-final` (see below)
Exemplos:	`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 *variáveis-independentes* 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 *Manual de Utilização do Gretl* for details.

Various internal variables may be retrieved using the genr command, provided genr is invoked immediately after this command.

The specific formula used for generating robust standard errors (when the `--robust` option is given) can be adjusted via the set command.

Caminho de Menu: /Model/Ordinary Least Squares

Acesso alternativo: Beta-hat button on toolbar

omit

Argumento:	*lista-de-variáveis*
Opções:	`--wald` (do a Wald test rather than an F-test)
	`--vcv` (mostrar matriz de covariância)
	`--quiet` (don't print estimates for reduced model)
	`--silent` (don't print anything)
	`--inst` (omit as instrument, TSLS only)
	`--both` (omit as both regressor and instrument, TSLS only)
Exemplos:	`omit 5 7 9`
	`omit seasonals --quiet`

This command must follow an estimation command. It calculates a test statistic for the joint significance of the variables in *lista-de-variáveis*, which should be a subset of the independent variables in the model previously estimated.

If the original model was estimated by OLS, the test statistic is by default an F-value. This is based on the sums of squared residuals for the restricted and unrestricted models, unless the original model was estimated with robust standard errors. In the latter case F is computed from the robust estimate of the covariance matrix for the original model. (It is the F-form of a Wald test).

For estimators other than OLS, or if the `--wald` option is given, the statistic is an asymptotic Wald chi-square value based on the covariance matrix of the original model.

By default, the restricted model is estimated, the estimates are printed, and the restricted model replaces the original as the "current model" for the purposes of, for example, retrieving the residuals as `$uhat` (or doing further tests such as add or omit).

If the `--wald` option is selected, the restricted model is not estimated (and so the current model is not replaced). The `--quiet` option suppresses the printout of the restricted model (if applicable): only the result of the test is printed. If the restricted model is both estimated and printed, the `--vcv` option has the effect of printing the covariance matrix for the coefficients in the restricted model, otherwise this option is ignored.

If the `--silent` option is given, nothing is printed; nonetheless, the results of the test can be retrieved using the special variables `$test` and `$pvalue`.

If the original model was estimated using two-stage least squares, an ambiguity arises: should the selected variables be omitted as regressors, as instruments, or as both? This is resolved as follows: by default the variables are dropped from the list of regressors, but if the `--inst` flag is given they are dropped as instruments, or if the `--both` flag is given they are dropped from the model altogether. These two options are incompatible with the `--wald` option; if one or more instruments are omitted the model must be re-estimated.

Caminho de Menu: Janela do modelo, /Testes/Omit variables

open

Argumento:	*ficheiro-de-dados*
Opção:	`--www` (use a database on the gretl server)
Exemplos:	`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. 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, etc.).

This command can also be used to open a database (gretl or RATS 4.0) 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.

Caminho de Menu: /Ficheiro/Open data

Acesso alternativo: Drag a data file into gretl (MS Windows or Gnome)

outfile

Argumentos:	*filename option*
Opções:	`--append` (append to file)
	`--close` (close file)
	`--write` (overwrite file)
Exemplos:	`outfile --write regress.txt`
	`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.

panel

Opções:	`--vcv` (mostrar matriz de covariância)
	`--random-effects` (random rather than fixed effects)
	`--between` (estimate the between-groups model)
	`--time-dummies` (include time dummy variables)
	`--unit-weights` (weighted least squares)
	`--iterate` (iterative estimation)
	`--quiet` (less verbose output)
	`--verbose` (more verbose output)

Estimates a panel model, by default using the fixed effects estimator. Depending on the number of cross-sectional units and the number of independent variables, this is implemented either by adding a set of dummy variables representing the units, or by subtracting the group or unit means from the original data.

If the `--random-effects` flag is given, random effects estimates are computed, using the method of 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).

For more details on panel estimation, please see *Manual de Utilização do Gretl*.

Caminho de Menu: /Model/Panel

pca

Argumento:	*lista-de-variáveis*
Opções:	`--save` (Save major components)
	`--save-all` (Save all components)

Principal Components Analysis. Prints the eigenvalues of the correlation matrix for the variables in *lista-de-variáveis* along with the proportion of the joint variance accounted for by each component. Also prints the corresponding eigenvectors (or "component loadings").

If the `--save` flag is given, components with eigenvalues greater than 1.0 are saved to the dataset as variables, with names `PC1`, `PC2` and so on. These artificial variables are formed as the sum of (component loading) times (standardized `Xi`), where `Xi` denotes the ith variable in *lista-de-variáveis*.

If the `--save-all` flag is given, all of the components are saved as described above.

Caminho de Menu: /View/Principal components

Acesso alternativo: Main window pop-up (multiple selection)

pergm

Argumentos:	*nome-de-variável* `[` *bandwidth* `]`
Opções:	`--bartlett` (use Bartlett lag window)
	`--log` (use log scale)

Computes and displays (and if not in batch mode, graphs) the spectrum of the specified variable. By default the sample periodogram is given; with the `--bartlett` flag 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.

When the sample periodogram is printed, two tests for fractional integration of the series ("long memory") are given, namely the Geweke and Porter-Hudak (GPH) test and the Local Whittle Estimator. The null hypothesis in both cases is that the integration order is zero. By default the order for these tests is the lesser of T/2 and T0.6. Again, this value can be adjusted using the bandwidth parameter.

Caminho de Menu: /Variável/Spectrum

Acesso alternativo: Menu de contexto da janela principal (selecção singular)

poisson

Argumentos:	*variável-dependente variáveis-independentes* `[ ;` *offset* `]`
Opções:	`--vcv` (mostrar matriz de covariância)
	`--verbose` (mostrar detalhes das iterações)
Exemplos:	`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,.... 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.

Caminho de Menu: /Model/Nonlinear models/Poisson

plot

Argumento:	*lista-de-variáveis*
Opção:	`--one-scale` (force a single scale)

Plots the values for specified variables, for the range of observations currently in effect, using ASCII symbols. Each line stands for an observation and the values are plotted horizontally. By default the variables are scaled appropriately. See also gnuplot.

print

Argumentos:	*lista-de-variáveis* ou *string-literal*
Opções:	`--byobs` (by observations)
	`--long` (use 10 significant digits or more)
	`--no-dates` (use simple observation numbers)
Exemplos:	`print x1 x2 --byobs`
	`print "This is a string"`

If *lista-de-variáveis* is given, prints the values of the specified variables; if no list is given, prints the values of all variables in the current data file. If the `--byobs` flag is given the data are printed by observation, otherwise they are printed by variable.

If the `--long` flag is given the data are printed, by variable, to greater than usual precision. The default in this case is to show 10 significant digits but you can adjust that figure using the set command.

If the `--byobs` flag is given and the data are printed 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.

If the argument to print is a literal string (which must start with a double-quote, `"`), the string is printed as is. See also printf.

Note: a special "hack" is available with this command, in conjunction with the `--byobs` flag, which can be useful when working with missing values in a data set. If you give a list of variables followed by a semi-colon, followed by one final variable, then the final variable is not printed but is used to screen the observations to print. Any observations for which the screening variable has value 0 are not printed. As an example of use, suppose you have a daily time series `x`, and you want a list of the dates for which `x` is missing. You can do

```
genr filt = missing(x)
print x ; filt --byobs
```

Caminho de Menu: /Data/Display values

printf

Argumentos:	*format args*

Prints scalar values and/or strings under the control of a format string (providing a small 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 in the dataset, (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
genr b = $coeff(2)
genr 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)
```

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 `%%`.

probit

Argumentos:	*variável-dependente variáveis-independentes*
Opções:	`--robust` (robust standard errors)
	`--vcv` (mostrar matriz de covariância)
	`--verbose` (mostrar detalhes das iterações)

If the dependent variable is a binary variable (all values are 0 or 1) maximum likelihood estimates of the coefficients on *variáveis-independentes* are obtained via the "binary response model regression" (BRMR) method outlined by Davidson and MacKinnon (2004). As the model is nonlinear the slopes depend on the values of the independent variables: the reported slopes are evaluated at the means of those variables. 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 Davidson and MacKinnon (2004, Chapter 10) for details.

If the dependent variable is not binary, but is discrete and has a minimum value of 0, then Ordered Probit estimates are obtained. In this case robust standard errors are not yet available. (If the variable selected as dependent is not discrete, or does not have a minimum of zero, an error is flagged.)

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

Caminho de Menu: /Model/Nonlinear models/Probit

pvalue

Argumentos:	*dist* [*params*] *xval*
Exemplos:	`pvalue z zscore`
	`pvalue t 25 3.0`
	`pvalue X 3 5.6`
	`pvalue F 4 58 fval`
	`pvalue G xbar varx x`
	`pvalue B bprob 10 6`

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, or `B` for binomial).

For the t and chi-square distributions the degrees of freedom must be given; for F numerator and denominator degrees of freedom are required; for gamma the mean and variance are needed; and for the binomial distribution the "success" probability and the number of trials must be given. In each case, these extra values are provided before the *xval*.

As shown in the examples above, the numerical parameters may be given in numeric form or as the names of variables.

Caminho de Menu: /Tools/P-value finder

pwe

Argumentos:	*variável-dependente variáveis-independentes*
Opção:	`--vcv` (mostrar matriz de covariância)
Exemplo:	`pwe 1 0 2 4 6 7`

Computes parameter estimates using the Prais–Winsten procedure, an implementation of feasible GLS which is designed to handle first-order autocorrelation of the error term. The procedure is iterated, as with corc; the difference is that while Cochrane–Orcutt discards the first observation, Prais–Winsten makes use of it. See, for example, Chapter 13 of Greene's *Econometric Analysis* (2000) for details.

Caminho de Menu: /Modelo/Série temporal/Prais-Winsten

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's *Introduction to Econometrics* (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).

Caminho de Menu: Janela do modelo, /Testes/Teste QLR test

quit

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

Caminho de Menu: /Ficheiro/Exit

rename

Variantes:	`rename` *varnumber newname*
	`rename` *varname newname*

Changes the name of the variable with identification number *varnumber* or current name *varname* 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.

Caminho de Menu: /Variável/Edit attributes

Acesso alternativo: Menu de contexto da janela principal (selecção singular)

reset

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 the cube of the fitted values to the regression and calculating the F statistic for the null hypothesis that the parameters on the two added terms are zero.

Caminho de Menu: Janela do modelo, /Testes/Ramsey's RESET

restrict

Imposes a set of linear restrictions on either (a) the model last estimated or (b) a system of equations previously defined and named. The syntax and effects of the command differ slightly in the two cases.

In both 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 implicitly to be applied to the last model, and they are evaluated as soon as the `restrict` command is terminated. In the system case the initial "restrict" must be followed by the name of a previously defined system of equations (see system). The restrictions are evaluated when the system is next estimated, using the estimate command.

Each restriction in the set should be expressed as an equation, with a linear combination of parameters on the left and a numeric value to the right of the equals sign. 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 equation may be prefixed with a numeric multiplier, using `*` to represent multiplication, 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 evaluated via a Wald F-test, using the coefficient covariance matrix of the model in question. By default, the restricted coefficient estimates are printed; if you just want the test statistic, you can append the `--quiet` option flag to the initial `restrict` command.

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.

Caminho de Menu: Janela do modelo, /Testes/Linear restrictions

rhodiff

Argumentos:	*rholist* ; *lista-de-variáveis*
Exemplos:	`rhodiff .65 ; 2 3 4`
	`rhodiff r1 r2 ; x1 x2 x3`

Creates rho-differenced counterparts of the variables (given by number or by name) in *lista-de-variáveis* and adds them to the data set, using the suffix `#` for the new variables. Given variable `v1` in *lista-de-variáveis*, and entries `r1` and `r2` in *rholist*, the new variable

```
v1# = v1 - r1*v1(-1) - r2*v1(-2)
```

is created. The *rholist* entries can be given as numerical values or as the names of variables previously defined.

rmplot

Argumento: *nome-de-variável*

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.

Caminho de Menu: /Variável/Range-mean graph

run

Argumento: *inputfile*

Execute the commands in *inputfile* 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.

Caminho de Menu: Run icon in script window

runs

Argumento: *nome-de-variável*

Carries out the nonparametric "runs" test for randomness of the specified variable. 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:

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

Caminho de Menu: /Variável/Runs test

scatters

Argumentos:	*yvar* `;` *xvarlist* ou *yvarlist ; xvar*
Opção:	`--with-lines` (create line graphs)
Exemplos:	`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 *xvarlist*, or of all the variables in *yvarlist* 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 six; 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.

Caminho de Menu: /View/Multiple graphs

sdiff

Argumento: *lista-de-variáveis*

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

Caminho de Menu: /Add/Seasonal differences of selected variables

set

Argumentos: *variável value*

Exemplos:
```
set qr on
set csv_delim tab
set horizon 10
```

Set the values of various program parameters. The given value remains in force for the duration of the gretl session unless it is changed by a further call to set. The parameters that can be set in this way are enumerated below. Note that the settings of `hac_lag` and `hc_version` are used when the `--robust` option is given to the ols command.

If the set command is given without any parameters, the current settings for all the relevant variables are printed.

- `echo`: `off` or `on` (the default). Suppress or resume the echoing of commands in gretl's output.
- `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.
- `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.
- `qr`: `on` or `off` (the default). Use QR rather than Cholesky decomposition in calculating OLS estimates.
- `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.
- `hac_lag`: `nw1` (the default) or `nw2`, or an integer. Sets the maximum lag value, 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, $p = 4 \times (T/100)^{2/9}$.
- `hc_version`: 0 (the default), 1, 2 or 3. Sets the variant used when calculating Heteroskedasticity Consistent standard errors with cross-sectional data. The 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".
- `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 (which does not take autocorrelation into account).
- `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.

- `hp_lambda`: `auto` (the default), or a numerical value. Sets the smoothing parameter for the Hodrick–Prescott filter (see the `hpfilt` function under the `genr` command). The default is to use 100 times the square of the periodicity, which gives 100 for annual data, 1600 for quarterly data, and so on.

- `bkbp_limits`: two integers, the second greater than the first (the defaults are 8 and 32). Sets the frequency bounds for the Baxter–King bandpass filter (see the `bkfilt` function under the `genr` command).

- `bkbp_k`: one integer (the default is 8). Sets the approximation order for the Baxter–King bandpass filter.

- `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.

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

- `bhhh_maxiter`: one integer, 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.

- `longdigits`: one positive integer value, less than or equal to 20. Determines the number of digits of precision used when printing the values of variables using the `--long` option (see print).

- `initvals`: a pre-specified matrix. Allows manual setting of the initial parameter estimates for ARMA estimation. For details see *Manual de Utilização do Gretl*.

setinfo

Argumentos:	*nome-de-variável* `-d` *description* `-n` *displayname*
Opções:	`--discrete` (mark variable as discrete)
	`--continuous` (mark variable as continuous)
Exemplos:	`setinfo x1 -d "Description of x1-n "Graph name"`
	`setinfo z --discrete`

Sets up to three attributes of the named variable, as follows.

If the `-d` 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 `-n` flag is given followed by a quoted string, that string is used to set the variable's "display name", which is then 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 variables as continuous; setting a variable as discrete affects the way the variable is handled in frequency plots.

Caminho de Menu: /Variável/Edit attributes

Acesso alternativo: Menu de contexto da janela principal

setobs

Variantes: setobs *periodicity startobs*
setobs *unitvar timevar*

Opções: `--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))

Exemplos:
```
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
```

Force the program to interpret the current data set as having a specified structure.

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.

Caminho de Menu: /Data/Dataset structure

setmiss

Argumentos: *value* [*lista-de-variáveis*]

Exemplos:
```
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`.

Caminho de Menu: /Sample/Set missing value code

shell

Argumento: *shellcommand*

Exemplos:
```
! 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

Variantes:	`smpl` *startobs endobs*
	`smpl` *+i -j*
	`smpl` *dumvar* `--dummy`
	`smpl` *condition* `--restrict`
	`smpl --no-missing [` *lista-de-variáveis* `]`
	`smpl` *n* `--random`
	`smpl full`
Exemplos:	`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 *lista-de-variáveis* is specified observations are selected on condition that all variables in *lista-de-variáveis* have valid values at that observation; otherwise, if no *lista-de-variáveis* is given, observations are selected on condition that *all* variables have valid (non-missing) values.

With the `--random` flag, the specified number of cases are selected from the full dataset at random. 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.

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`: Any "structural" information in the data file (regarding the time series or panel nature of the data) is lost when this command is issued. You may reimpose structure with the setobs command.

Please see *Manual de Utilização do Gretl* for further details.

Caminho de Menu: /Sample

spearman

Argumentos: *x y*
Opção: `--verbose` (print ranked data)

Prints Spearman's rank correlation coefficient for the two variables x and y. 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 first. For example:

```
genr altx = -x
spearman altx y
```

Caminho de Menu: /Model/Robust estimation/Rank correlation

sprintf

Argumentos: *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*.

square

Argumento: *lista-de-variáveis*
Opção: `--cross` (generate cross-products as well as squares)

Generates new variables which are squares of the variables in *lista-de-variáveis* (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.

Caminho de Menu: /Add/Squares of selected variables

store

Argumentos: *ficheiro-de-dados* [*lista-de-variáveis*]
Opções: `--csv` (use CSV format)
`--omit-obs` (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)
`--database` (use gretl database format)
`--overwrite` (see below, on database format)

Saves either the entire dataset or, if a *lista-de-variáveis* is supplied, a specified subset of the variables in the current dataset, to the file given by *datafile*.

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 *Manual de Utilização do Gretl* for details on the various formats.

The option flag `--omit-obs` is 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; only the actual data are printed.

Note that any scalar variables will not be saved automatically: if you wish to save scalars you must explicitly list them in *lista-de-variáveis*.

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.

Caminho de Menu: /Ficheiro/Save data; /File/Export data

string

Argumentos:	*stringvar* = *expression*
Exemplos:	`string mystr = "P-value:  "`
	`string homedir = getenv("HOME")`

Saves a string named *stringvar*. The *expression* can take the following forms: a literal string, enclosed in double quotes; a series of such strings, separated by spaces (these strings will be concatenated); or a call to the `getenv` function to retrieve a string from the program's environment.

After defining a string variable in this way, the variable can be printed using its name preceded by `@`, as in

```
print "mystr = @mystr"
```

See also printf, sprintf.

summary

Argumento:	[*lista-de-variáveis*]

Print summary statistics for the variables in *lista-de-variáveis*, or for all the variables in the data set if *lista-de-variáveis* is omitted. Output consists of the mean, standard deviation (sd), coefficient of variation (= sd/mean), median, minimum, maximum, skewness coefficient, and excess kurtosis.

Caminho de Menu: /View/Summary statistics

Acesso alternativo: Menu de contexto da janela principal

system

Variantes:	`system method=`*estimator*
	`system name=`*sysname*
Argumento:	*savevars*
Exemplos:	`system name="Klein Model 1"`
	`system method=sur`
	`system method=sur save=resids`
	`system method=3sls save=resids,fitted`
	Ver também`klein.inp`, `kmenta.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 give 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.

In the optional save= field of the command you can specify whether to save the residuals (resids) and/or the fitted values (fitted).

For examples of the specification and estimation of systems of equations, please see the scripts `klein.inp`, `kmenta.inp` and `greene14_2.inp` (supplied with the gretl distribution).

Caminho de Menu: /Model/Simultaneous equations

tabprint

Argumento:	`[` *-f filename* `]`
Opções:	`--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 the form of a LaTeX table. 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`, where `N` is the number of models estimated to date in the current session. See also eqnprint.

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`.

Caminho de Menu: Janela do modelo, /LaTeX

testuhat

Must follow a model estimation command. Gives the frequency distribution for the residual from the model along with a chi-square test for normality, based on the procedure suggested by Doornik and Hansen (1984).

Caminho de Menu: Janela do modelo, /Testes/Normality of residual

tobit

Argumentos:	*variável-dependente variáveis-independentes*
Opções:	`--vcv` (mostrar matriz de covariância)
	`--verbose` (mostrar detalhes das iterações)

Estimates a Tobit model. This model 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. For details see Greene's *Econometric Analysis*, Chapter 20.

Caminho de Menu: /Model/Nonlinear models/Tobit

tsls

Argumentos:	*variável-dependente variáveis-independentes* ; *instruments*
Opções:	`--vcv` (mostrar matriz de covariância)
	`--robust` (robust standard errors)
Exemplo:	`tsls y1 0 y2 y3 x1 x2 ; 0 x1 x2 x3 x4 x5 x6`

Computes two-stage least squares (TSLS or IV) estimates: *depvar* is the dependent variable, *variáveis-independentes* is the list of independent variables (including right-hand side endogenous variables) in the structural equation for which TSLS estimates are needed; and *instruments* is the combined list of exogenous and predetermined variables in all the equations. If the *instruments* list is not at least as long as *variáveis-independentes*, the model is not identified.

In the above example, the `y`s are the endogenous variables and the `x`s are the exogenous and predetermined variables.

Output 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 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. Davidson and MacKinnon (2004, chapter 8) give a good explanation of both tests.

Caminho de Menu: /Model/Other linear models/Two-Stage least Squares

var

Argumentos:	*ordem lista-de-variáveis* `[ ;` *exolist* `]`
Opções:	`--nc` (não incluir uma constante)
	`--seasonals` (incluir variáveis sazonais auxiliares)
	`--robust` (robust standard errors)
	`--impulse-responses` (print impulse responses)
	`--variance-decomp` (print forecast variance decompositions)
	`--lagselect` (show information criteria for lag selection)
Exemplos:	`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 *lista-de-variáveis* — they will be added automatically. The semi-colon separates the stochastic variables, for which *order* lags will be included, from any exogenous variables in *exolist*. 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 run for each variable in *lista-de-variáveis*. 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.

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.

Caminho de Menu: /Modelo/Série temporal/Vector autoregression

varlist

Prints a listing of variables currently available. list and ls are synonyms.

vartest

Argumentos:	*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.

Caminho de Menu: /Model/Bivariate tests/Difference of variances

vecm

Argumentos:	*ordem rank lista-de-variáveis*
Opções:	`--nc` (sem constante)
	`--rc` (constante restringida)
	`--crt` (constante e tendência restringida)
	`--ct` (constante e tendência não restringida)
	`--seasonals` (incluir auxiliares sazonais centradas)
	`--impulse-responses` (print impulse responses)
	`--variance-decomp` (print forecast variance decompositions)
Exemplos:	`vecm 4 1 Y1 Y2 Y3`
	`vecm 3 2 Y1 Y2 Y3 --rc`
	Ver também`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 *lista-de-variáveis*.

lista-de-variáveis 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 Johensen (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 *Manual de Utilização do Gretl*.

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.

Caminho de Menu: /Modelo/Série temporal/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, Wasserman, and Kutner (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.

Caminho de Menu: Janela do modelo, /Testes/Collinearity

wls

Argumentos:	*wtvar variável-dependente variáveis-independentes*
Opções:	`--vcv` (mostrar matriz de covariância)
	`--robust` (robust standard errors)
	`--quiet` (não mostrar resultados da regressão)

Computes weighted least squares (WLS) estimates using *wtvar* as the weight, *depvar* as the dependent variable, and *variáveis-independentes* 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* `*` *variáveis-independentes*. The *R*-squared, however, is calculated in a special manner, namely as

$$R^2 = 1 - \frac{\mathrm{ESS}}{\mathrm{WTSS}}$$

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*.

Caminho de Menu: /Model/Other linear models/Weighted Least Squares

xcorrgm

Argumentos:	*var1 var2* `[` *maxlag* `]`
Exemplo:	`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 a *maxlag* 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.

Caminho de Menu: /View/Cross-correlogram

Acesso alternativo: Menu de contexto da janela principal (multiple selection)

xtab

Argumentos:	*ylist* `;` *xlist*
Opções:	`--row` (display row percentages)
	`--column` (display column percentages)
	`--zeros` (display zero entries)

With no options given, displays a contingency table or cross-tabulation for each variable in *ylist* (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.

The `--row` and `--column` options are mutually exclusive, and instead of the frequency count yield the percentages for each row or column, respectively. The `--zeros` option 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.

Capítulo 1

Reserved Words

Reserved words, which cannot be used as the names of variables, fall into the following categories:

- Names of constants: `CONST`, `NA`, `const`, `null`, `pi`.
- Names of internal variables and data types: `matrix`, `obs`, `scalar`, `series`, `t`.
- Names of functions, as shown in Table 1.1.

Tabela 1.1: Function names

BFGSmax	I	abs	acos	argname	asin	atan	bkfilt
cdemean	cdf	cdiv	ceil	cholesky	cmult	cnorm	colnames
cols	corr	cos	cov	critical	cum	det	diag
diff	dnorm	dsort	dummify	eigengen	eigensym	exp	fdjac
fft	ffti	filter	firstobs	floor	fracdiff	gammafun	genpois
getenv	gini	ginv	grab	hpfilt	imaxc	imaxr	iminc
iminr	infnorm	int	inv	invcdf	invpd	islist	isnull
isscalar	isseries	isstring	lags	lastobs	ldet	ldiff	lincomb
ljungbox	ln	lngamma	log	log10	log2	lower	lrvar
makemask	max	maxc	maxr	mcorr	mcov	mean	meanc
meanr	median	mexp	min	minc	minr	missing	misszero
mlag	mnormal	mols	movavg	mread	mshape	msortby	muniform
mwrite	mxtab	nelem	nobs	normal	nullspace	obslabel	obsnum
ok	onenorm	ones	orthdev	pdf	pmax	pmean	pmin
pnobs	polroots	princomp	psd	pvalue	qform	qnorm	qrdecomp
quantile	randgen	rank	ranking	rcond	readfile	resample	round
rows	sd	sdc	sdiff	selifc	selifr	seq	sin
sort	sortby	sqrt	sst	strlen	strstr	sum	sumc
sumr	svd	tan	tr	transp	trimr	uniform	unvech
upper	values	var	varname	varnum	vec	vech	wmean
wsd	wvar	xpx	zeromiss	zeros			