Gauss for Econometrics: Sample Programs

R.G. Pierse

In representing examples of *GAUSS* code in these notes, the following conventions are adopted: all variables are in *italic* typeface and all *GAUSS* reserved words are in **bold** typeface. Roman typeface is used to represent strings and the names of procedures other than *GAUSS* intrinsic functions.

1. OLS Regression

The procedure **olsq** computes OLS regression. It takes three arguments: an $n \times 1$ vector of observations on the regressand y, an $n \times k$ matrix of observations on the regressors X, and a scalar, $nfcst \geq 0$, defining how many observations are to be reserved for post-sample forecast tests. The procedure returns four arguments: the $k \times 1$ vector of coefficient values b, the $k \times 1$ vector of coefficient standard errors seb, the $k \times k$ variance-covariance matrix, and a 10×2 matrix of regression diagnostics, D. For each diagnostic, the first column gives the value of the test statistic, and the second column gives (where appropriate) the p-value from the correct distribution. The diagnostics are: $\overline{R}^2$, $\widehat{\sigma}^2$, the Durbin-Watson DW statistic, the log-likelihood function LLF, the Godfrey (1978) LM test for 4th order serial correlation, the Ramsey (1969) RESET test for functional misspecification, the Jarque and Bera (1980) test for normality, the LM test for heteroscedasticity as used in MicroFit, and, if $nfcst > 0$, the Chow (1960) test for predictive failure and the Chow test for coefficient stability.

```
proc (4) = olsq(y,X,nfcst);
/* (c) Richard G. Pierse 1996 */
local  n1,n2,n3,n,m,k,df,D,b,XX,Ul,s2,seb,b0,rss0,b2,rss2,
       u,y2,u2,y2m,u2m,mu1,mu2,mu3,mu4,rsq;
n1 = 1; n3 = rows(y); n2 = n3-nfcst; n = n2-n1+1;
```

```
k = cols(X); df = n-k;
D = zeros(10,2);
/* XX = (X'X)^-1 */
XX=invpd(X[n1:n2, . ]'X[n1:n2, . ]);
/* b = the vector of coefficient estimates: (X'X)^-1 X'y */
b = XX*X[n1:n2, . ]'y[n1:n2];
/* u = the vector of OLS residuals: y - Xb */
u = y[n1:n2]-X[n1:n2, . ]*b;
/* s2 = the equation standard error: σ̂^2 */
s2 = (u'u)/df ;
/* seb = the vector of coefficient standard errors */
seb = sqrt(s2*diag(XX));
if nfcst > 0;
  b0 = invpd(X[n1:n3, . ]'X[n1:n3, . ])*X[n1:n3, .]'y[n1:n3];
  rss0 = sumc( (y[n1:n3] - X[n1:n3, .]*b0)^2) ;
  if nfcst > k;
    b2 = invpd(X[n2+1:n3, . ]'X[n2+1:n3, . ])
           *X[n2+1:n3, . ]'y[n2+1:n3];
    rss2 = sumc( (y[n2+1:n3]-X[n2+1:n3, . ]*b2)^2);
  endif;
endif;
y = y[n1:n2]; X = X[n1:n2, . ];
y2 = (y - u)^2; u2 = u^2;
y2m = y2 - meanc(y2); u2m = u2 - meanc(u2);
mu1 = meanc(u); mu2 = (u'u)/n; mu3 = (u'u2)/n; mu4 = (u2'u2)/n;
/* R-Squared R^2 */
rsq = 1-(u'u)/sumc((y-meanc(y))^2);
/* R-Bar Squared R̄^2 */
```

```
D[1,1] = (rsq-(k-1)/df*(1.0-rsq));
/* Equation standard error σ̂ */
D[2,1] = sqrt(s2);
/* Durbin-Watson statistic */
D[3,1] = (sumc((u[2:n]-u[1:n-1])^2)/(u'u));
/* log-likelihood function */
D[4,1] = -n/2*(1+ln(2*pi*mu2));
/* Ul is matrix of lagged residuals = [u_-1 : u_-2 : u_-3 : u_-4] */
Ul = shiftr((u.*.ones(1,4))',seqa(1,1,4),0)';
/* Godfrey (1978) test for 4th order serial correlation */
D[5,1] = n*(u'Ul*invpd(Ul'Ul-Ul'X*XX*X'Ul)*Ul'u)/(u'u);
D[5,2] = cdfchic(D[5,1],4) ;
/* Ramsey (1969) RESET test of functional form */
D[6,1] = n*((u'y2)/(y2'y2-y2'X*XX*X'y2)*(y2'u)/(u'u));
D[6,2] = cdfchic(D[6,1],1) ;
/* Jarque-Bera (1980) test for normality */
D[7,1] = n*((mu3^2)/(6*mu2^3)+(((mu4/(mu2^2))-3)^2)/24
        +3*(mu1^2)/(2*mu2) - (mu3*mu1)/(mu2^2));
D[7,2] = cdfchic(D[7,1],2) ;
/* MicroFit test for heteroscedasticity */
D[8,1] = n*((u2m'y2m)^2)/(y2m'y2m)/(u2m'u2m);
D[8,2] = cdfchic(D[8,1],1) ;
 if nfcst > 0;
   /* Chow (1960) Predictive failure test */
   D[9,1] = (rss0-u'u)/s2;
   D[9,2] = cdfchic(D[9,1],nfcst) ;
   if nfcst > k;
      /* Coefficient stability test */
```

```
    D[10,1] = ((rss0-u'u-rss2)/(u'u+rss2))*(n+nfcst-2*k);
    D[10,2] = cdfchic(D[10,1],k) ;
  endif;
endif;
retp (b,seb,s2*XX,D);
endp;
```

2. ADF Tests

Procedure **adf** computes augmented Dickey-Fuller tests for unit roots (Dickey and Fuller (1979, 1981) and Fuller (1976)). It takes three arguments: an $n \times 1$ vector of observations on the variable y, a scalar $k \geq 0$, defining the order of augmentation, and a scalar nc, determining how many deterministic components to include in the regression. If $nc = 0$, then no intercept or time trend is included. If $nc = 1$, an intercept but no trend is included. If $nc = 2$, both intercept and trend are included. The procedure has two scalar returns: the *ADF* statistic *df*, and the 95% significance level for the statistic, *cv*. This confidence level is computed using the response surfaces calculated by MacKinnon (1991).

```
proc (2) = adf(y,k,nc);
/* (c) Richard G. Pierse 1996 */
local n,dy,yl,X,XX,b,s2,df,cv,cvs;
/* MacKinnon (1991) response surface coefficients */
cvs[3,3] = {-1.9393 -0.398 0, -2.8621 -2.738 -8.36,
                -3.4126 -4.039 -17.83};
n = rows(y);
dy = y[2:n]-y[1:n-1];
yl = y[1:n-1];
n = n-1;
/* X = [y_{-1} : Δy_{-1} : · · · : Δy_{-k} : ι : t] */
X = yl;
```

```
if k > 0; X=X~shiftr(ones(k,1)*dy',seqa(1,1,k),zeros(k,1))'; endif;
X = X~ones(n,1)~seqa(1,1,n);
X = X[k+1:n ,1: k+nc+1]; dy = dy[k+1:n];
XX = invpd(X'X);
b = XX*X'dy;
s2 = sumc((dy-X*b)^2)/(n-2*k-nc-1);
df = b[1]/sqrt(s2*XX[1,1]);
cv = cvs[nc+1,1]+cvs[nc+1,2]/(n-k)+cvs[nc+1,3]/((n-k)^2);
retp (df,cv);
endp;
```

3. Johansen Procedure

Procedure **coint** implements the Johansen (1988, 1991) procedure for testing for cointegration in a multivariate framework.This is based on the maximum likelihood estimation of the *VECM* model

$$\Delta \mathbf{y}_t = \sum_{i=1}^{p-1} \mathbf{A}_i \Delta \mathbf{y}_{t-i} + \boldsymbol{\beta}\, \boldsymbol{\alpha}' \mathbf{y}_{t-p} + \boldsymbol{\mu} + \boldsymbol{\delta}' \mathbf{x}_t + \mathbf{u}_t$$

where the *VAR* model can include an intercept term $\boldsymbol{\mu}$ and a set of $I(0)$ exogenous variables $\mathbf{x}_t$.

The log-likelihood function of this model can be written as

$$L(\boldsymbol{\alpha}) = c - \frac{T}{2} \sum_{i=1}^{p} \log(1 - \lambda_i)$$

where λ_i are the solutions to the *generalised eigenvalue* problem

$$\left| \boldsymbol{\lambda} \mathbf{S}_{kk} - \mathbf{S}_{k0} \mathbf{S}_{00}^{-1} \mathbf{S}_{0k} \right| = 0 \qquad (3.1)$$

where $\mathbf{S}_{ij} = T^{-1} \sum_{t=1}^{T} \mathbf{R}_{it} \mathbf{R}'_{jt}\, , i, j = 0, k,$ and $\mathbf{R}_{0t}$ and $\mathbf{R}_{kt}$ are the vectors of residuals from regressing $\Delta \mathbf{y}_t$ and $\mathbf{y}_{t-p}$ respectively, on $\{\Delta \mathbf{y}_{t-1}, \cdots, \Delta \mathbf{y}_{t-p+1}, \boldsymbol{\mu}, \mathbf{x}_t\}$. The number of cointegrating vectors, r, is equal to the number of *non-zero* eigenvalues, λ_i.

There is no procedure in *GAUSS* for solving the generalised eigenvalue problem (3.1). However, since $\mathbf{S}_{kk}$ is symmetric and positive-definite, then

$$\mathbf{S}_{kk}^{-1} = \mathbf{H}'\mathbf{H}$$

where $\mathbf{H}$ is the upper triangular Cholesky decomposition of $\mathbf{S}_{kk}^{-1}$. It follows that

$$\left|\boldsymbol{\lambda}\mathbf{S}_{kk} - \mathbf{S}_{k0}\mathbf{S}_{00}^{-1}\mathbf{S}_{0k}\right| = \left|\boldsymbol{\lambda}\mathbf{I}_{k} - \mathbf{H}\mathbf{S}_{k0}\mathbf{S}_{00}^{-1}\mathbf{S}_{0k}\mathbf{H}'\right| = 0$$

which is a *standard* eigenvalue problem that can be solved in *GAUSS* using the **eighv** procedure.

coint takes four arguments: an $T \times k$ matrix of observations on the k variables Y, an $T \times n_{i0}$ matrix of observations on the $I(0)$ exogenous variables X, (or zero if there are no exogenous variables), a scalar parameter *lag* to signify the lag length of the *VAR* process (p), and a scalar parameter *model* to determine the deterministic components to include in the model. The permissible values of model are:

0 no intercept: $\boldsymbol{\mu} = 0$

1 restricted intercept (intercept only in error correction term)

$$\Delta\mathbf{y}_t = \sum_{i=1}^{p-1} \mathbf{A}_i \Delta\mathbf{y}_{t-i} + \boldsymbol{\beta}\left(\boldsymbol{\alpha}'\mathbf{y}_{t-p} + \boldsymbol{\alpha}_0\right) + \boldsymbol{\delta}'\mathbf{x}_t + \mathbf{u}_t$$

where $\boldsymbol{\beta}\boldsymbol{\alpha}_0 = \boldsymbol{\mu}$.

2 unrestricted intercept

3 True data process has a restricted intercept but the model with an unrestricted intercept is estimated.

There are seven returns from the procedure **coint**:

1. the $k \times 1$ vector of eigenvalues $\boldsymbol{\lambda}$
2. the $k \times k$ matrix of cointegrating vectors $\boldsymbol{\alpha}$
3. the $k \times k$ matrix of coefficients on the cointegrating vectors $\boldsymbol{\beta}$
4. the $T \times k$ matrix of the error correction terms $\mathbf{Y}\boldsymbol{\alpha}$

5. a $k \times 2$ matrix *maxstat*, presenting results from the Johansen *maximal eigenvalue tests*. The ith row of the first column of *maxstat* presents a test statistic for there being exactly $i-0$ cointegrating vectors against an alternative of more than $i-1$. The corresponding element in the second column presents the 95% critical value for this test as derived from Osterwald-Lenun (1992) and Johansen (1995)

6. a $k \times 2$ matrix *trstat*, presenting results from the Johansen *trace tests*. The ith row of the first column of *trstat* presents a test statistic for there being exactly $i-0$ cointegrating vectors against an alternative of more than $i-1$. The corresponding element in the second column presents the 95% critical value for this test as derived from Osterwald-Lenun (1992) and Johansen (1995)

7. a $(k+1) \times 4$ matrix *crit*, presenting model selection criteria for the Johansen models. The ith row refers to the model with $i-0$ cointegrating vectors. The first column is the value of the maximised log-likelihood function, the second column is the Akaike Information Criterion, the third column is the Schwartz Bayesian criterion and the fourth column is the Hannan-Quinn criterion. These criteria can be used to select the correct lag length of the *VAR* and to test between the models *0-3*, but only once the number of cointegrating vectors r has been determined on the basis of the *maxstat* or *trstat* tests.

```
proc (7) = coint(Y,X,lag,model);
/* (c) Richard G. Pierse 1996 */
local t,n,k,ni0,DY,Z,ZZ,R0,Rk,S00,S01,S0k,Skk,
    lambda,alpha,beta,ecm,maxstat,trstat,
    cvmax,cvtrace,llf,npar,crit;
/* 95% critical values from Osterwald-Lenun (1992) */
cvmax={
        3.84 11.44 17.89 23.80 30.04
        36.36 41.51 47.99 53.69 59.06,
        9.243 15.672 22.002 28.138 34.400
        40.303 46.455 51.995 57.422 63.571,
```

```
        3.762 14.069 20.967 27.067 33.461
        39.372 45.277 51.420 57.121 62.805,
        8.176 14.900 21.074 27.136 33.319
        39.426 44.912 51.071 56.996 62.419};
cvtrace={
        3.84 12.53 24.31 39.89 59.46
        82.49 109.99 141.20 175.77 212.67,
        9.243 19.964 34.910 53.116 76.069
        102.139 131.700 165.579 202.920 244.148,
        3.762 15.410 29.680 47.210 68.524
        94.155 124.243 155.999 192.887 233.135,
        8.176 17.953 31.525 48.280 70.598
        95.177 124.253 157.109 192.840 232.486};
n=rows(Y); k=cols(Y); ni0=cols(X);
if rows(X) ne n; ni0=0; endif;
t=n-lag;
DY=Y-(zeros(1,k)|Y[1:n-1,.]);
Z=zeros(n,1); if model gt 1; Z=Z~ones(n,1); endif;
if lag gt 1;
  Z=Z~shiftr((ones(1,lag-1).*.DY)',
  seqa(1,1,lag-1).*.ones(k,1),0)';
endif;
if ni0 gt 0; Z=Z~X; endif;
DY=DY[lag+1:n,.];
R0=DY;
Rk=Y[1:t,.]; if model eq 1; Rk=Rk~ones(t,1); endif;
if cols(Z) gt 1;
  Z=Z[lag+1:n,2:cols(Z)];
```

```
  ZZ=invpd(Z'Z);
  R0=DY-Z*ZZ*Z'DY;
  Rk=Y[1:t,.]-Z*ZZ*Z'Y[1:t,.];
  if model eq 1;
    Rk=Rk~(ones(t,1)-Z*ZZ*Z'ones(t,1));
  endif;
endif;
S00=R0'R0/t; S0k=R0'Rk/t;
Skk=(invpd(Rk'Rk/t));
S01=S00-S0k*Skk*S0k';
S00=invpd(S00);
Skk=chol(Skk);
llf=-t/2*(k*ln(2*pi)+k+ln(det(S01)));
/* Call eigenvalue procedure */
{lambda,alpha}=eighv(Skk*S0k'S00*S0k*Skk');
/* Order eigenvalues in decreasing order */
lambda=rev(lambda); alpha=rev(alpha')';
/* Normalise α and β */
alpha=Skk'alpha/sqrt(t); beta=S0k*alpha*t;
/* Calculate ECM terms */
ecm=Y; if model eq 1; ecm=ecm~ones(n,1); endif;
ecm=ecm*(alpha./(ones(rows(alpha),1)*alpha[1,.]));
/* Maximal eigenvalue statistic */
maxstat=zeros(k,2);
maxstat[.,1]=-t*ln(1-lambda[1:k]);
maxstat[.,2]=rev(cvmax[model+1,1:k]');
/* Trace statistic */
trstat=zeros(k,2);
```

```
trstat[.,1]=upmat(ones(k,k))*maxstat[.,1];
trstat[.,2]=rev(cvtrace[model+1,1:k]');
/* Model selection criteria */
crit=zeros(k+1,4);
/* Model log-likelihood */
crit[.,1]=llf-((trstat[.,1]/2)|0);
npar=(k*(k*lag+ni0)-(k-seqa(0,1,k+1))^2)/2;
if model eq 1;
   npar=npar+(k-seqa(0,1,k+1))/2;
elseif model eq 2;
   npar=npar+k/2;
endif;
/* Akaike Information Criterion */
crit[.,2]=crit[.,1]-2*npar;
/* Schwartz Bayesian Criterion */
crit[.,3]=crit[.,1]-npar*ln(t);
/* Hannan-Quinn Criterion */
crit[.,4]=crit[.,1]-2*npar*ln(ln(t));
retp(lambda,alpha,beta,ecm,maxstat,trstat,crit);
endp;
```

References

[1] Chow, G.C. (1960), 'Tests of equality between subsets of coefficients in two linear regression models', *Econometrica*, 28, 591–605.

[2] Dickey, D.A. and W.A. Fuller (1979), 'Distribution of the estimators for autoregressive time series with a unit root', *Journal of the American Statistical Association*, 74, 427–431.

[3] Dickey, D.A. and W.A. Fuller (1981), 'Likelihood ratio statistics for autoregressive time series with a unit root', *Econometrica*, 49, 1057–1072

[4] Engle, R.F. and C.W.J. Granger (1991), *Long Run Economic Relationships: Readings in Cointegration*, Oxford University Press, Oxford.

[5] Fuller, W.A. (1976), *Introduction to Statistical Time Series*, Wiley, New York.

[6] Godfrey, L.G. (1978), 'Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables', *Econometrica*, 46, 1293–1302.

[7] Jarque, C.M. and A.K. Bera (1980), 'Efficient tests for normality, homoscedasticity and serial independence of regression residuals', *Economic Letters*, 6, 255–9.

[8] Johansen, S. (1988), 'Statistical analysis of cointegrating vectors', *Journal of Economic Dynamics and Control*, 12, 231–54.

[9] Johansen, S. (1991), 'Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models', *Econometrica*, 59, 1551–80.

[10] Johansen, S. (1995), *Likelihood-based Inference in Cointegrated Vector Autoregressive Models*, Oxford University Press, Oxford.

[11] MacKinnon, J.G. (1991), 'Critical values of cointegration tests', in R.F. Engle and C.W.J. Granger (eds.) *Long Run Economic Relationships: Readings in Cointegration*, Oxford University Press, Oxford.

[12] Osterwald-Lenum, M. (1992), 'A note with fractiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics: four cases', *Oxford Bulletin of Economics and Statistics*, 54, 461–472.

[13] Ramsey, J.B. (1969), 'Tests for specification errors in classical least squares regression analysis', *Journal of Royal Statistical Society, B*, 31, 350–71.