R E G - X

by

Stephen G. Hall

USER GUIDE

R E G - X

ROWNEY FARM,CHASEWAYS,SAWBRIDGEWORTH,HERTS,CM21 OAS,UK

----

version 98.1

REG-X is a general econometric estimation package written by Stephen Hall for the IBM PC and compatibles (XT/AT/386), it is interactive and relatively easy to use while offering a range of advanced econometric techniques.

SINGLE VARIABLE ANALYSIS

Simple descriptive statistics, tests of stationarity and normality and spectral analysis.

BIVARIATE ANALYSIS

Parametric and nonparametric tests of the distribution of two data sets and their correlation

REGRESSION ANALYSIS

Extensive OLS and IV estimation for dynamic time series models with a wide range of diagnostics and procedures including; Recursive estimation, CUSUM test, CUSUMSQ test, Sequential CHOW test, LM for serial correlation, RESET test, ARCH test, Normality test, Dickey-Fuller tests for cointegration, correlogram, Box-Pierce test, Ljung-Box test, Hendry Forecast test, correlogram of squared residuals. Johansen maximum likelihood estimation procedure for multiple cointegrating vectors with likelihood ratio tests. Engle and Yoo 3-step estimation.

TIME VARYING PARAMETER MODELS

Single equation models may be estimated using the Kalman Filter with stochastic parameters which follow a random walk and which include a stochastic trend term in the equation. Backward smoothing and forecasting are also offered as part of this option.

GARCH-M ESTIMATION

Single equation GARCH-M and SGARCH-M models may be estimated by maximum likelihood.

MATRIX ANALYSIS

Matrices may be entered and manipulated interactively using a range of options, eg. Matrix multiplication, inversion, transpose, addition, calculation of determinant, Cholesky decomposition, calculation of eigenvalues and vectors.

SEASONAL MODELS AND GENERAL KALMAN FILTER

A new general model of seasonality is implemented. Very general models in State Space form may be estimated by maximum likelihood.

****

HARDWARE REQUIREMENTS: an IBM PC or compatible with a hard disk, 16Mb of RAM and RUNNING WINDOWS95 or WINDOWSNT. REG-X is supplied on 3.5" high density floppy disks unless some other format is requested. PRICE: Commercial licence £400.00. Academic site licence £50. Payment by non-sterling cheque please add £10 to the license fee.

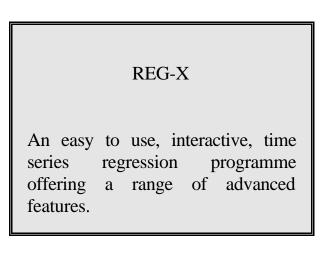**Please make all cheques payable to S.G.Hall.**

CONTENTS

1. INTRODUCTION

REG-X is an interactive program, designed for the IBM PC and compatibles, for the analysis of econometric times series data and the estimation of econometric models. It offers a range of sophisticated techniques in a form which is easily accessible to a user who is familiar with the underlying principles of dynamic economic modelling. A user

REG-X

An easy to use, interactive, time series regression programme offering a range of advanced features.

who understands the econometric principles involved should be able to use the program with little prior instruction, section 2 outlines the program installation procedure and some basic conventions after reading this such a user is recommended to try the program and refer back to the rest of the manual only as needed.

Many of the underlying econometric principals of REG-X are discussed in **CUTHBERTSON, HALL and TAYLOR, `APPLIED ECONOMETRIC TECHNIQUES' published 1991 by Phillip Allen**, this manual will discuss only the specific implementation of these ideas in REG-X and the user is referred to this book for background detail.

2. INSTALLING REG-X AND SOME BASIC CONVENTIONS

2.1   INSTALLATION

REG-X is supplied on a single high density floppy disk, it will run under windows95/98 and windows NT.

To install REG-X insert the disk in the floppy disk drive and from the programme manager run the INSTALL programme.   Follow the instructions and the program will be installed onto the hard disk of the machine.

A programme group will be created with an icon for REG-X and an unistall programme to remove REG-X. To run REG-X, simply click on the REG-X icon.

**REG-X
INSTALLATION**

this shows the easy to follow steps for installing REG-X.

Working directory

REG-X expects to be able to write to a sub directory called

C:\temp

This directory should be created if one does not already exist on the C: drive

In a network installation where there is no write access to C: then the default directory may be changed by the following procedure.

In the REG-X installation directory there is a text file called default.dir. this contains one line, the location of the default scratch directory. This may be changed to any drive or location where REG-X can have write access.

2.2 CONVENTIONS

REG-X uses a range of conventions,most of which are set up in the initial few responses after running the program. these are :-

Frequency

The frequency of the data is simply the periodicity of the data ie. weekly, quarterly, annual, monthly etc. It is defined by a single number which is defined as annual(1), quarterly(4), monthly(12), weekly(52), and undated(100).

Dates

All dates are referred to using a four or six digit code where the first two or four digits refer to the year and the last two refer to the sub period within the year. If we are using quarterly data then 8801 (or 198801) is the first quarter of 1988, annual data must be referred to with a 1 in the last place so 1986 is 8601. Undated data may be referred to as 0001. A date such as 6001 will be taken to be 196001 so any dates beyond 1999 must be entered as six digits.

Base period

REG-X will ask for the earliest period which is to be used in the run, this is the base period. It will align all data from that period onwards and it will not use data before that period. Many data files can be read into REG-X one after another and they do not have to be over the same data period, the base period is used to align these files consistently. For example, if the base period is 6501 then we could read in a data file which ran from 6701 to 7501, REG-X would then pad out the first two years with missing values, we could then read in another file with data over the period 6201 to 7801, REG-X would then drop the data before 6501 and correctly align the remaining data with the first data set.

Typing characters and names

REG-X will distinguish between upper case letters and lower case letters, a command such as ADD must be in upper case as in the examples provided on the screen. Variable names must be in the same case as when they were originally defined, so if you name a variable cpi REG-X will not recognize CPI as the same variable. In general it is best to work solely in upper case

characters, this will avoid confusion and no responses are ever required in lower case.

Output

When you start REG-X you will be asked for the name of an output file, any file name may be given as a suitable response, from that point on a constant record of any procedures performed and results obtained will be kept in the output file. This file may later be edited or printed. In an extended REG-X session the output file can become very long and so there is an option within REG-X to open a new output file. This closes the existing file and starts a new one. If the new file has the same name as the old one the first output file will be erased.

2.3 TEST DATA FILE

REG-X is supplied with a test data file called REG-X.DAT, this file is in sequential format with a base year of 6501

3. DATA MANIPULATION

3.1 INTRODUCTION
REG-X can read data from a range of formats it can graph individual data series, it can perform data transformations and it can write out new data files and tables of data. All these functions are performed through the data manipulation option (option 1) from the main menu.

3.2 DATA INPUT

Data may be read in from a file in one of the following 4 formats

By Variable

This is an ascii (human readable, you can edit this file) file where all the variables are over the same period and are arranged in blocks of numbers so that each block is the data for one variable. So all the data for the first variable comes in a block first then all the data for the second variable and so on. When reading a file by variable REG-X will ask you for the file name, the dates that the data runs over,the number of variables in the file and the names of the variables in the order that they occur in the file.

By Observation

This is an ascii file where all the variables are over the same period and are arranged in a single block by observation. So first comes the data for all the variables for the first observation and then comes the data for all the variables for the next observation and so on. When reading a file by observation REG-X will ask you for the file name, the dates that the data runs over,the number of variables in the file and the names of the variables in the order that they occur in the file.

> **REG-X**
> **DATA MANIPULATION**
>
> REG-X is not primarily designed as a vehicle for data manipulation so complex data transformations may often be done more easily in another programme. However you should be able to do anything you need to here as well as graphing and printing the data.

Sequential

This is an ascii file which is in a special REG-X format so that it includes the names and dates for the data inside the file. When saving a data file for future use you should always save it in sequential format as this is the quickest format to read back in.

EID format

This format has only one variable in each file with a header card it is designed so individual variables may be transferred between REG-X runs or data files.

Entered by hand

Data may be typed in by hand directly into the program, data already in REG-X may be changed by hand.

Data from spreadsheets

Data from a spreadsheet can be introduced into REG-X by creating a data file from the spreadsheet either by variable or by observation in ascii format. All spreadsheets are capable of doing this, in symphony it is done by using the PRINT option where you print to a designated file rather than direct to the printer.

3.3 CREATING NEW DATA FILES

After performing transformations or saving results it is often useful to save the data file so that it may be used for future work. This is done using the option to Write out data files. When you chose this option you will be able to save data either by variable, by observation, in EID format or in sequential format. If you want to use the data for future REG-X work sequential is invariably the easiest format to use. REG-X always saves these files with a set name of REG-X.DAT REG-X.VAR or REG-X.OBS and it will over write any existing files with those names.

3.4 TABULATING DATA

The print data to output file option will either produce a full listing of the data on the output file or tabulate subsets of the variables both on the screen and in the output file.

3.5 GRAPHING VARIABLES

The graph a variable option will produce a time series graph of any of the variables in the data set. The graphs may be printed to either an EPSON configured printer or on a Hewlett Packard laser Jet II configured printer by choosing the appropriate option. NOTE: printing on a laser printer takes time and a pause of several minutes may occur before any output appears.

3.6 DATA TRANSFORMATIONS

Within the data manipulation there is an option to carry out data transformations and create trends,seasonal and dummy variables.

4.   SINGLE   VARIABLE ANALYSIS

4.1 INTRODUCTION

Single variable analysis provides a range of descriptive statistics for a single variable, these are aimed primarily at describing the time series properties of the variable.

4.2   LIST   OF   DESCRIPTIVE STATISTICS

REG-X
SINGLE VARIABLE ANALYSIS

This option offers a range of single variable descriptive statistics, stationarity tests, spectrum, simple statistics, correlogram, etc.

The following statistics are provided automatically, they are calculated for any chosen period;

Sum of the variable
Mean of the variable
Standard error of the variable
Autocorrelation function for the variable up to the   16th order
Box-Pierce test for a random white noise series up to   16th order
Ljung-Box test for a random white noise series up to   16th order
Dickey-Fuller unit root test
Augmented Dickey-Fuller unit root test (with four   lagged delta terms)
Phillips test for stationarity
Coefficient of skewness
Coefficient of Kurtosis
Bera-Jarque test for normality
The Spectrum of the series
A graph of the spectrum

5. BIVARIATE ANALYSIS

5.1 INTRODUCTION

Bivariate analysis provides two sets of test procedures, the first tests two series to see if they come from the same distribution (it could be the same series over different time periods) and the second tests two series to see if they are correlated. Both procedures offer a range of parametric and non-parametric tests.

REG-X
BIVARIATE ANALYSIS

This option provides a range of parametric and non-parametric measures of the correlation between two series

5.2 TESTING TWO SERIES FOR EQUAL DISTRIBUTION

Two series are provided to this procedure, they do not have to be over the same data period and indeed they may even be the same series over different data periods. The following tests are then provided to test whether the series have the same underlying distribution.

STUDENT T TEST OF DIFFERENCES IN MEAN(with the assumption of the same variance)

this is calculated as

$$T = (XM_1 - XM_2)/S_d$$

where XM_i is the mean of series i, and

$$S_d = ((1/N_1 + 1/N_2)(\Sigma(X_{i1} - XM_1)^2 + \Sigma(X_{i2} - XM_2))/(N_1 + N_2 - 2))^{0.5}$$

that is S_d is the average standard deviation of both variables. The probability level shows the probability that T is this large purely by chance.

STUDENT T TEST OF DIFFERENCES IN MEANS(with the assumption of different variances)

this is calculated as

$$T = (XM_1 - XM_2)/(Var(X_1)/N_1 + Var(X_2)/N_2)^{0.5}$$

where Var(X) is the variance of X. The implied probability shows the probability that the T value could have arisen by chance.

STUDENT T TEST OF DIFFERENCES IN MEANS FOR PAIRED DATA

This test is appropriate where the data comes in pairs which may have been affected by some third factor, so we are observing a much larger variance than we might otherwise have seen if this third factor had been held constant. We generally expect this to be a more powerful test than the two above, where it is appropriate. This test is calculated as:

$$T = (XM_1 - XM_2)/S_d$$

where

$$S_d = ((Var(X_1) + Var(X_2) - 2Cov(X_1, X_2))/N)^{0.5}$$

where N is the number of pairs.

F TEST OF DIFFERENCES IN VARIANCE

This test is simply calculated as the ratio of the variances of the two series, the implied probability is the probability that the Null (they do have the same variance) is true.

KOLMOGOROV-SMIRNOV TEST

This is a non parametric test which tests if the two series come from the same, unspecified distribution. REG-X reports the D statistic, D is a function of the deference between the cumulative density functions of the two series. The

implied probability shows the probability that the Null hypothesis (that the distributions are the same) is true.

5.3 TESTING THE CORRELATION BETWEEN TWO SERIES

This part of REG-X provides three tests of the correlation between two series over some common data period.

LINEAR CORRELATION ANALYSIS

The most widely used measure of association between two series is the correlation coefficient. This is defined as

$$r = \Sigma(X_1-XM_1)(X_2-XM_2)/((\Sigma(X_1-XM_1)^2)^{0.5}(\Sigma(X_2-XM_2)^2)^{0.5})$$

The implied probability is the probability that their is zero correlation between the two series.

SPEARMAN RANK-ORDER PROCEDURE

This procedure is a non-parametric one which operates using the notion of ranks, that is we replace each element of X_1 and X_2 with its rank in the series where the smallest element is 1 and the next smallest is 2 and so on then if there is a high correlation between the two series we expect the ranks to closely match. This transforms the problem from one of correlation between series with unknown distributions to two series with a known distribution - namely uniform distributed over 1 to N. We can then construct tests for the correlation which exists between these two series. The implied probabilities show the probability that there is no correlation between the two series.

KENDALL'S TAU
This procedure is another non-parametric technique which is based on ranks but in this case rather than using the numerical rank it simply uses the relative ordering of the ranks of the two series, ie higher or lower in rank. Kendall's tau is usually believed to be more robust than the two procedures outlined above but it may often be less powerful at detecting a correlation which actually exists.

6. REGRESSION ANALYSIS

6.1 INTRODUCTION

REG-X has extensive facilities for single equation OLS and Instrumental Variable (IV) estimation, it performs a large range of diagnostics including recursive estimation. The program has also been developed with the new Cointegration techniques particularly in mind and it can be easily used to implement the Engle and Granger two stage estimation procedure as well as the Maximum likelihood estimation of multiple cointegrating vectors suggested by Johansen. REG-X will also produce simple analysis of the step responses of a complex dynamic equation to shocks in the exogenous variables.

> REG-X
> REGRESSION ANALYSIS
>
> This option performs OLS and IV regression analysis with a wide range of summary statistics and diagnostics. It also offers the Johanson method of analysing cointegration between a set of variables and an option to investigate the dynamic response of an equation to shocks.

The modelling principles which REG-X employs is extensively discussed in Cuthbertson, Hall and Taylor chapter 4.

7. TIME VARYING PARAMETER MODELS

REG-X will estimate linear single equation models with time varying parameters of the following general form.

> **REG-X**
> **TIME VARYING MODELS**
>
> This option estimates models with time varying parameters using the Kalman Filter, it also allows the models to include a stochastic trend term.

$$Y_t = T_t + B_t X_t + u_t$$

where B_t is a vector of time varying parameters which are generated from the following process

$$B_t = B_{t-1} + v_t$$

T_t is a stochastic trend term which is generated by the following process

$$T_t = T_{t-1} + C_{t-1} + w_t$$

$$C_t = C_{t-1} + z_t$$

u,v,w and z are error terms with zero mean.

The model then contains a trend term which can vary both in terms of its rate of growth and in terms of its level in a stochastic way, as well as a set of parameters which follow a random walk process.

This model will be estimated using the Kalman Filter, this requires a set of hyper-parameters which show the relative variance between the main structural equation error term (called the measurement equation) u_t and the error terms on the equation governing the evolution of the trend and parameters v_t, w_t and z_t called the state equations. REG-X will first estimate these hyper-parameters by maximizing the concentrated likelihood function, it will then provide estimates of

the trend and parameters as produced by the Kalman Filter, it then goes on to estimate a smoothed version of the model. At this point REG-X offers the option to modify the hyper-parameters by scaling them, this can produce either more or less erratic parameter estimates. Finally REG-X will use the Kalman Filter model to forecast beyond the end of the estimation period, note that if there are stochastic parameters this can only be done if there is data for the exogenous variables beyond the final estimation date. A pure stochastic trend model can of course be used to forecast without any future data.

The Kalman Filter is discussed in Cuthbertson, Hall and Taylor, chapter 7.

8.MATRIX ANALYSIS

REG-X offers a range of matrix manipulation routines under the matrix manipulation option. Matrices may be added, multiplied, transposed, inverted, the determinate may be calculated, a cholesky decomposition may be performed, eigenvalues may be calculated and they may also be read from or written to a file.

> REG-X
> MATRIX MANIPULATION
>
> This option offers a range of standard matrix calculations, determinants, eigenvalues and eigenvectors, etc.

The matrices which are used during the calculation of the Johansen maximum likelihood procedure are automatically stored in a file called JOHANSEN.MAT which may be read directly into the matrix manipulation routines.

9. THE GARCH-M AND SGARCH-M MODEL

9.1 INTRODUCTION

This section (based on Hall 1990) shows how the stochastic GARCH-M (SGARCH-M) model may be put into state space form and estimated by the Kalman Filter, it shows that conventional GARCH-M estimation is a special case of SGARCH-M

estimation when the GARCH equation is non-stochastic and there is no uncertainty about the initial conditions. The Kalman Filter provides a useful way of relaxing these two implausible assumptions.

9.2 THE GARCH-M MODEL

The basic GARCH-M model consists of

$$y_t = \mathbf{a}' x_t + \mathbf{d} h_t^2 + \mathbf{e}_t \qquad \textbf{1}$$

$$\mathbf{e}_t / \Omega_{t-1} \sim N(0, h_t^2)$$

$$h_t^2 = \mathbf{g}_0 + \sum_{i=1}^{n} \mathbf{g}_i \mathbf{e}_{t-i}^2 \qquad (2)$$

where x_t is a vector of conditioning variables, weakly exogenous for all parameters of (1) and (2). Engle Lilien and Robins(1987) term this kind of model ARCH-in-mean or ARCH-M. Note the h_t^2 is the conditional variance of ε_t formed at period t based on the information set (σ-field) of all information up to period t-1, and by assumption h_t^2 is non-stochastic.

A further extension of the ARCH formulation, which imposes smoother behaviour on the conditional second moments, has been suggested by Bollerslev

(1986). In Bollerslev's GARCH formulation, the conditional second moments are functions of their own lagged values as well as the squares and cross-products of lagged forecast errors. Bollerslev did not consider the GARCH-M extension although this is a fairly obvious one which was subsequently used in Bollerslev, Engle and Wooldridge(1988). Thus, for example, the GARCH-M (n, p) formulation of the above model would consist of (1) and

$$h_t^2 \;=\; A_0 \;+\; \sum_{i=1}^{n} A_i \, e_{t-i}^2 \;+\; \sum_{i=1}^{p} B_i \, h_{t-i}^2 \qquad\qquad \mathbf{3}$$

where the B_i and A_i are coefficients.

Stacking all of the parameters of the system into a single vector

$$\mathbf{m} = (\, \boldsymbol{a}, \boldsymbol{d}, A_0 \,, A_1 ,..., A_n , B_1 ,..., B_p \,) \qquad\qquad \mathbf{4}$$

and applying Schweppe's (1965) prediction error decomposition form of the likelihood function, the log-likelihood for a sample of T observations (conditional on initial values) is proportional to

$$L(\mathbf{m}) \;=\; \sum_{t=1}^{T} (\text{-}\log|h_t^2| \;-\; e_t^2 \,/\, h_t^2 \,) \qquad\qquad \mathbf{5}$$

(where we have assumed normality of the forecast errors).

Although the analytic derivatives of (5) can be computed (see Engle, Lilien and Robins, 1987) variable-metric algorithms which employ numerical derivatives are simpler to use and easily allow changes in specification.   Under suitable regularity conditions (Crowder, 1976), maximization of (5) will yield maximum likelihood estimates with the usual properties.

The difficulty with the GARCH equation (3) when it is not exact can be most easily appreciated by focusing on the information sets which are being used. Assume that h^2_0 and ε_0 are known and that n=p=1 then

$$h_1^2/\Omega_0 \;=\; A_0 \;+\; A_1\,\boldsymbol{e}_0^2 + B_1 h_0^2 \qquad\qquad \textbf{6}$$

so h_1^2 is properly conditioned on information at period 0. But when we use equation (3) to form an estimate of h_2^2 this is no longer the case, to begin with,

$$\boldsymbol{e}_1 \;=\; y_1 \;-\; \boldsymbol{a}\,x_1 - \boldsymbol{d}(\,h_1^2/\Omega_0\,) \qquad\qquad \textbf{7}$$

So ε_1 is based on a mixture of information, most of which is dated at t=1 but the conditional variance is dated at t=0. The main question asked in this paper is; shouldn't we be able to improve our estimate of h^2 once full information for period 1 is available over and above our initial estimate of h_1^2 formed only on information known in period 0. The normal procedure would be to use the initial estimate in (3) when we come to period 2. Thus,

$$h_2^2 \;=\; A_0 \;+\; a_1\,\boldsymbol{e}_1^2 \;+\; B_1(\,h_1^2/\Omega_0\,) \qquad\qquad \textbf{8}$$

So the estimate of h_2^2 is based on Ω_0 and partly on y_1 and x_1. In order to fully utilize the information set Ω_1 we should update our estimate of h_1^2 when observations on y_1 and x_1 become available and also derive a new estimate of the error ε_1 based on Ω_1. These updated estimates should then be used in (3) in the second period.

To estimate the stochastic GARCH-M model we can cast the model into state space form and use the Kalman Filter. The GARCH-M model is already very close to a state space representation, all we need to do is to reinterpret some of the notation of the state space model. For simplicity I will only consider the GARCH(1,1) model but higher order GARCH models can be treated in an exactly analogous fashion. To begin with, in the state equation we interpret w_t to be a 2x1 vector which is made up of a constant and ε^2_{t-1},this equation is then identical to the GARCH equation (3) where we identify z_t with h_t^2. The z_t term in the measurement equation is then seen as the variance term which is identical to the variance term in equation (1), so we only need to make explicit allowance for the structure of the variance of the measurement equation and this is done simply

by equating Γ_t with z_t. The SGARCH-M state space model may then be written out as,

$$y_t = \boldsymbol{d}_{z_t} + \boldsymbol{a}' x_t + \boldsymbol{e}_t \tag{17}$$

$$\boldsymbol{e}_t \sim N(0, z_{t|t-1})$$

and

$$z_t = B_1 z_{t-1} + A_0 + A_1 \boldsymbol{e}_{t-1}^2 + \boldsymbol{w}_t$$

$$\tag{18}$$

$$\boldsymbol{w}_t \sim N(0, Q_t)$$

This model would be identical to the equations (1) and (3) when $Q_t=0$ and $P_0=0$, that is when the GARCH equation is non stochastic with certain initial conditions. This can be seen by noting that if $Q_t=0$ and $P_0=0$, then the updating equations are non operational and $f_t=\Gamma_t=z_t$ and the likelihood function for the Kalaman Filter becomes equivalent to (5).

When this is not the case the Kalman Filter equations may be used to estimate the model for known values of the parameters of the model. When any of the parameters or variances are unknown this procedure may be used to define the likelihood function conditional on the unknown parameters and hence may be used for maximum likelihood estimation of the stochastic GARCH-M model.

The use of the standard Kalman Filter does however raise one potential problem with regard to the sign of the conditional variance z_t. Under the assumption that P_0 is non zero and Q_t is normally distributed it is always possible that z_t could be updated to a negative value. In practice this can be seen as extremely unlikely for sufficiently small values of P_0 and Q_t, but it is certainly an important conceptual problem. One solution is of course to assume that ψ_t follows some distribution other that the normal one, the log-normal distribution would be a suitable choice, an extended Kalman Filter algorithm could then be defined to cope with this assumption. The standard Kalman Filter also presents certain practical difficulties as the use of diffuse priors in the form of a very large estimate of P_0

will often produce estimates of z_t which become negative in the early periods of the recursion, and so this attractive assumption is not fully operational. An extended Kalman Filter is however computationally unattractive and would of course still be dependent on the specific distributional assumption made. Another solution to the problem is to assume that P_0 follows a truncated normal distribution, thus the filter works in the usual way except when a negative value for z_t would result and then z_t would simply be set to some small (non zero value).   This of course can not be justified on the grounds of Maximum likelihood estimation but it may be viewed as a Quasi- or pseudo-maximum likelihood estimator. Thus under this assumption while the estimation will not be fully efficient it will be consistent. This is the estimation technique which will be used below in the example.

The use of the Quasi-Maximum Likelihood (QML) argument does however raise a further problem, the usual ML estimators for the covariance matrix are not consistent. There are two main ways of consistently estimating the covariance matrix of the parameters under maximum likelihood based on the following two matrices,

$$A(\boldsymbol{m}) \ = \ T^{-1} \sum_{t=1}^{T} \partial^2 L_t \ \ / \ \partial \boldsymbol{m}_i \partial \boldsymbol{m}_j \qquad (19)$$

$$B(\boldsymbol{m}) \ = \ T^{-1} \sum_{t=1}^{T} \partial L_t(\boldsymbol{m}) / \partial \boldsymbol{m} . \ \partial L_t(\boldsymbol{m}) / \partial \boldsymbol{m} \qquad (20)$$

where μ is the vector of parameters being estimated, $A(\mu)$ is the Hessian matrix and $B(\mu)$ is the score matrix and when the full ML assumptions hold the parameter estimates are asymptotically normal with the distribution,

$$\sqrt{T}(\boldsymbol{m}^{*} - \boldsymbol{m}) \_ N(0, A(\boldsymbol{m})^{-1}) \ = \ N(0, B(\boldsymbol{m})^{-1}) \qquad (21)$$

so that the inverses of both the hessian and the score matrices are asymptotically equivalent and both are valid estimates of the covariance matrix. Under QML estimation where the full ML assumptions do not hold this is not the case and A and B will not be asymptotically equivalent and neither will be a consistent estimator of the true covariance matrix. The QML estimate of μ, μ^* is asymptotically normal with the following distribution,

$$\sqrt{T}(\boldsymbol{\mu}^*-\boldsymbol{\mu})\_N(0,C(\boldsymbol{\mu})) \tag{22}$$

where $C(\boldsymbol{\mu})$ is defined as

$$C(\boldsymbol{\mu}) = A(\boldsymbol{\mu})^{-1}B(\boldsymbol{\mu})A(\boldsymbol{\mu})^{-1}$$

Clearly when A and B are equivalent (14) reduces to the standard estimate of either A or B but when this equivalence does not hold this is the appropriate estimate of the covariance matrix. (14) is the basis for all the various formulae for robust estimates of the covariance matrix which have been used over recent years.

References

Baba Y, Engle R F, Kraft D F and Kroner K F (1987). 'Multivariate Simultaneous Generalised ARCH' Mimeo University of California San Diego.

Bollerslev T, (1986). 'Generalised Autoregressive Conditional Heteroskedasticity', Journal of Econometrics, vol. 31 no.3 pp. 307-328

Bollerslev T, Engle R F and Wooldridge J M (1988). 'A Capital Asset Pricing Model with Time-varying Covariances', Journal of Political Economy,vol.96, no.1 pp.116-131.

Crowder M J, (1976). 'Maximum Likelihood Estimation for Dependent Observations', Journal of the Royal Statistical Society, Series B, vol.38, no.1 pp.45-53.

Engle R F, (1982). 'Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of UK Inflation' Econometrica, vol.55, no.4 pp.987-1008

Engle R F, (1987). `Multivariate GARCH with Factor Structures - Cointegration in Variance', paper presented at a conference on time-varying variances in money and finance, University of California at San Diego, April.

Engle R F, Lilien D and Robins R, (1987). 'Estimating Time Varying Risk Premia in the Term Structure: The ARCH-M Model', Econometrica, vol.55, no.2 pp.391-407.

Griliches Z(1983).'Introduction:Data and Econometrics- The Uneasy Alliance.' chapter 25, 'Handbook of Econometrics' vol. 33 (ed) Griliches z. and Intriligator M.D. North Holland

Harvey A C. (1987) `Applications of the Kalman Filter in Econometrics.` in Bewley T F.(ed) 'Advances in Econometrics. Fifth World Congress' vol 1, Econometric Society Monograph No.13 Cambridge University Press.

Schweppe F, (1965). 'Evaluation of Likelihood Functions for Gaussian Signals', IEEE Transactions on Information Theory, vol.11, no.1, pp.61-70.

Van Praag B.(1983) 'The Population-Sample Decomposition in Minimum Distance Estimation' Unpublished, Presented at the Harvard-MIT Econometric Seminar

White H, (1982). 'Maximum Likelihood Estimation of Misspecified Models', Econometrica Vol 50 no.1 pp.1-25.

10. GENERAL KALMAN FILTER

The state space form is an extremely flexible one which allows many models to be estimated conveniently through the use of the Kalman Filter. This option in REG-X allows an extremely general state space model to be set up and estimated. The generality is achieved by setting the state space matrices up in individual files on the default directory, the elements of these matrices may then be either fixed values, they may be determined by maximum likelihood or they may (in some cases) be assigned

> ### REG-X
> ### KALMAN FILTER
>
> This option offers a very general Kalman Filter procedure to estimate models which are specified in state space form. Elements of the matrices may be estimated by maximum likelihood, they may be fixed to known values or they may be set equal to a variable from the data matrix. When the estimation is complete the model may also be smoothed

the value of a particular variable in the data matrix. The general form of the model is

Measurement Equations

$$\begin{array}{ccccccc} Y & = & Z \, . \, A & + \, D \, . \, X & + & v & v{\sim}N(0,W) \\ nx1 & & nxm \ mx1 & nxq \ qx1 & & nx1 & nxn \end{array}$$

State, or Transition Equations

$$\begin{array}{ccccccc} A(t) & = & T \, . \ A(t{-}1) & + \ R \, . \, X & + & e & e{\sim}N(0,Q) \\ mx1 & & mxm \ mx1 & mxq \ qx1 & mx1 & & mxm \end{array}$$

where the dimension sizes are shown under each matrix, (dimension limitations are n=10, m=25, q=10).

The matrices are input as disk files with the name of each matrix and the extension .KAL, the full set are Z.KAL, D.KAL, X.KAL, W.KAL, T.KAL, R.KAL and Q.KAL

The problem is defined in a file called SET.KAL, this has the following form

Line 1

    start date   end date   (eg 7001 8601)

Line 2

   n m q   (eg 2 8 1)

Line 3   then n lines giving the names of the measurement variables.

    VAR1

    VAR2   (if n=2)

Each of the matrix files has the following form. First the numerical values for the complete matrix, then how many elements of the matrix are to be maximized, then the elements to be maximized (in terms of their matrix coordinates). And, in the case of the Z matrix and the X matrix, how many elements are to be set at the value taken by a particular variable, again followed by the elements to be set as variables and the variable name (which must be in REG-Xs data matrix). To illustrate the Z matrix for the case n=2 and m=3.

Z.KAL

1 2 3
5 3 1
2   parameters to be maximised
   1   1   (1 1 is the element in row 1 column 1)
   2   1   (2 1 is the element in row 2 column 1)
2   parameters to be set as variables
   1   3VAR1   (note these lines are in fixed format,
   2   3VAR2   variable names must start in column 9)

So for example, the first element of the first row of this matrix would be maximized, taking the value 1 as its starting value and the third element of the first row would be set to the value of VAR1 in each period.

Generally it will be easier to create and changes these files directly by using an editor but REG-X will set these files up for you by prompting you interactively for their specification. This may help if the precise format is proving difficult.

When you go into the General Kalman routine REG-X will ask if you have the necessary files set up, if you say `N' then it will help you to create them.

In a general unobserved components model of this form it can be very difficult numerically to find good estimates of both Q and W in an unrestricted form, all that really matters is there ratio to each other and the scaling is not easy to pin down. For this reason the Likelihood function is concentrated so that one element of Q or W which would normally be maximised must be set to 1. The output then gives a scaling factor which is the maximum likelihood estimate of this parameter, the value of all the other elements of Q or W should also be scaled by this factor.

NOTE: The Kalman Filter is an extremely general tool, if your problem doesn't work well it may not be a fault with the program but with the specification of your model. In particular it is very easy to specify un-identified models in this general set up, or to get your specification of the matrices wrong.