Jurgen A. Doornik, David F. Hendry
and Sébastien Laurent

An Introduction to

XlModeler™ 1

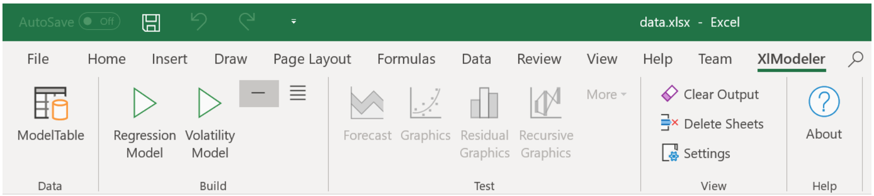
www.xlmodeler.com

Contents

Chapter 1

XlModeler ModelTable

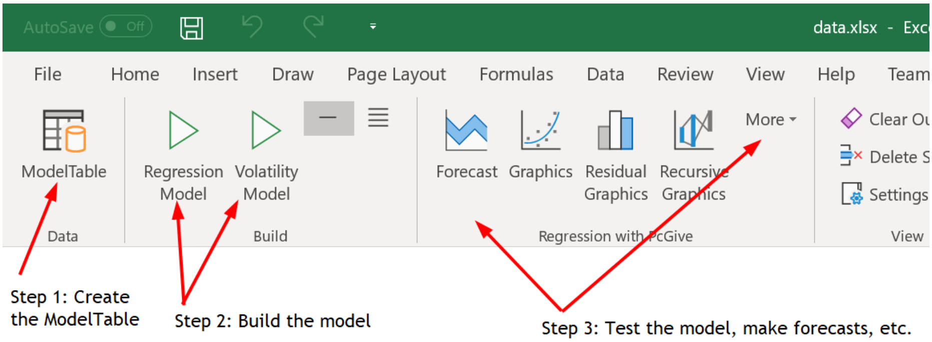
The functionality of XlModeler is available though its ribbon menu:



Using XlModeler consists of three steps:

1. Create the ModelTable,
2. Build a model,
3. Test the model.

After a model has been built successfully, the evaluation icons and dropdown menu come available:



To start using XlModeler, it is necessary to create a *ModelTable*. This defines the variables that can be used to build a model. The ModelTable can be saved with the workbook, so does not need to be redone everytime.

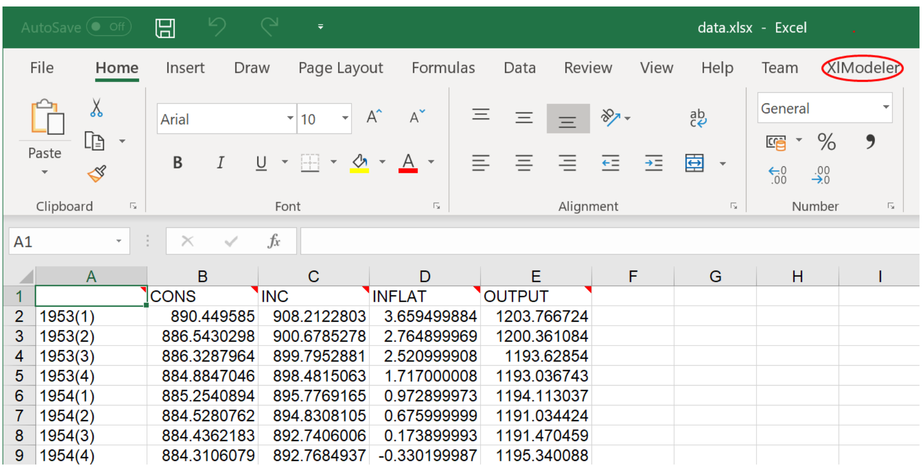
A simple artificial data set, `data.xlsx` is included with XlModeler, and installed in your user folder. We shall use this to illustrate the process.

Open the `data.xlsx` workbook. In this case we wish to add the entire datasheet to the *ModelTable*. There are two easy ways to do this:

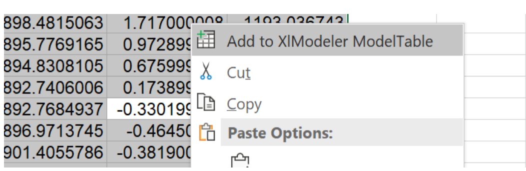
1. Select and right click,
2. Use the dialog, then select and add.

1.1 ModelTable from selection

The data set looks like this:



Select a cell and press `Ctrl+A`: this selects a rectangular part of the sheet around the cell, all relevant data in this case. Alternatively, click in the top left cell to select the entire sheet. Next, right click in the selection:



Select `Add to XlModeler ModelTable`. This creates a new worksheet, entitled **XlModeler.Table**, which records the selection:

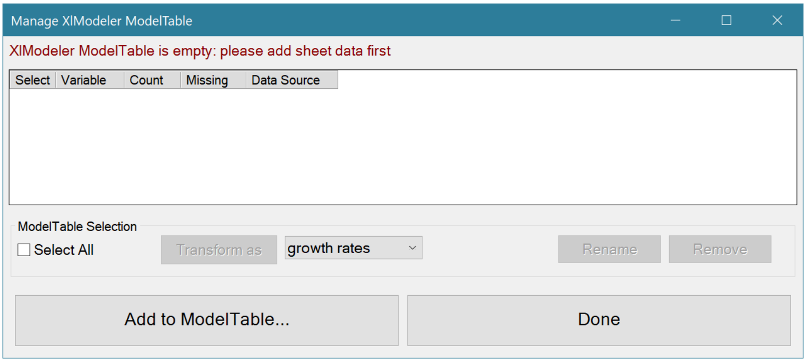
	A	B	C	D	E
1	**ModelTable**				
2	Frequency	4			
3	Start Year	1953			
4	Start Period	1			
5	Dates Style	Fixed Frequency			
6	Dates Source	1!A2:A160			
7					
8	**Variables**				
9	**Name**	**Data Source**	**Count**	**Missing**	
10	CONS	1!B2:B160	159	0	
11	INC	1!C2:C160	159	0	
12	INFLAT	1!D2:D160	159	0	
13	OUTPUT	1!E2:E160	159	0	
14	Constant	ones()	-1	0	
15	Trend	trend()	-1	0	
16	Seasonal	period()==1	-1	0	
17					

Four variables were created from sheet **1**. At the top we see that the ModelTable has a fixed frequency of 4 (quarterly data), starting in 1953(1).[1] Three additional variables were automatically created: a constant (for the regression intercept), trend, and seasonals (provided the seasonal frequency is not unity). These are available for regression models.

The sample information of the table is now determined, but further variables may be added. The actual observations are collected from the sheet just before estimating a model.

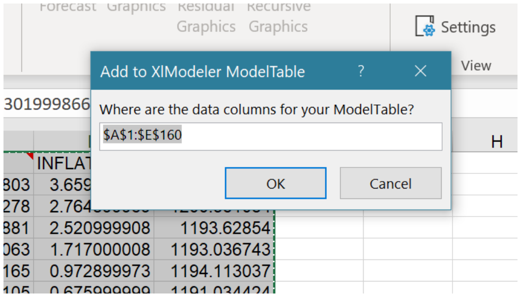
1.2 ModelTable dialog

The ModelTable dialog shows the current model, and allows variables to be renamed or removed. A limited choice of transformations can also be made. When it is empty:

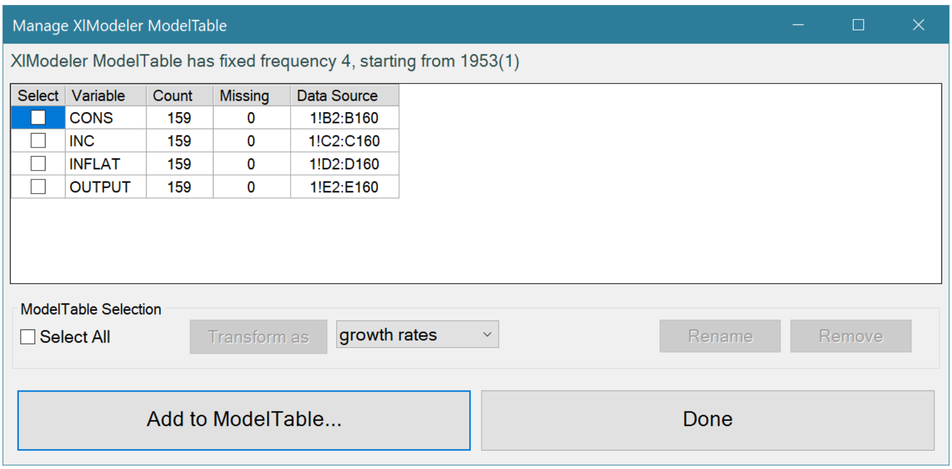


Now we can use the Add to ModelTable button to make a selection:

[1]With other data sets the seasonality may not be picked up correctly. In that case it can be modified here.



After which we have an updated dialog:



Pressing Done updates the **XlModeler.Table** sheet as shown above.

1.3 ModelTable transformations

The following transformations can be made directly in the ModelTable for a variable y_t with frequency S:
growth rate (%) $S 100 \Delta \log(y_t)$,
logarithm $\log(y_t)$,
difference Δy_t,
year-on-year growth (%) $100 \Delta_S \log(y_t)$,
year-on-year difference $\Delta_S y_t = y_t - y_{t-S}$.
Examples are given in §3.5 and Chapter 4.

Chapter 2

Regression Model

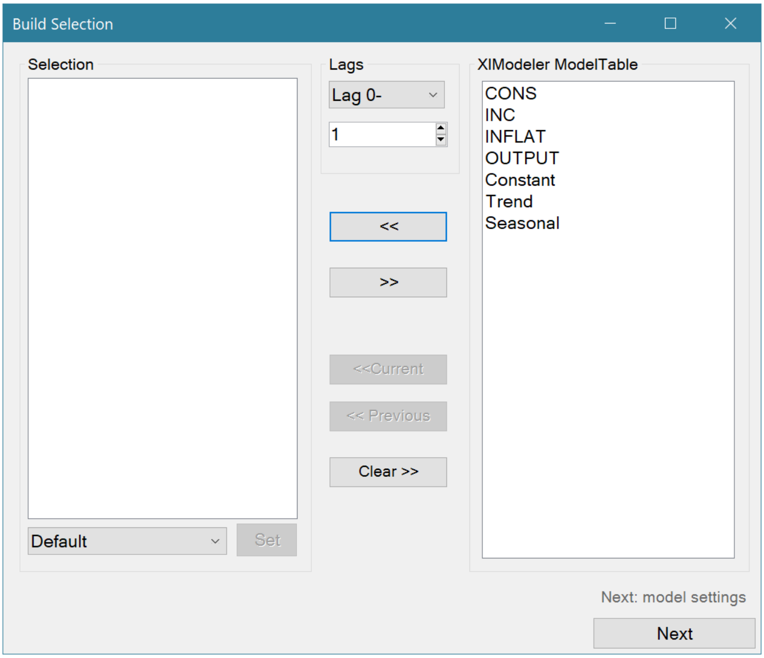
2.1 Introduction

Regression modelling normally consists of a cycle of two steps: build the model, evaluate the estimated model, then reformulate as required. Tis process takes place using the XlModeler toolbar in Excel. This tutorial will guide you through a simple model sequence based on the artificial data set. Hopefully you will agree at the end that XlModeler combines sophistication with great simplicity.

Before we start, a brief digression on lags is called for. XlModeler names lagged variables by appending an underscore and then the lag length. So `CONS_1` is `CONS` one period lagged. XlModeler uses this naming scheme to keep track of the lag length. Note that lags are never created in the ModelTable. Instead, they are created through the build process, allowing XlModeler to keep track of them.

2.2 Build a model, step 1: selection

Here we use the `data.xlsx` workbook and ModelTable as created in the previous chapter. Then click on the Regression Model icon to initiate the Build dialog. The dialog is grouped in three columns. On the right are the variables that can be added from the ModelTable to the selection (i.e. the model that we are building). In the middle are buttons for moving between the ModelTable and the selection. For dynamic models, the lag length is also there. On the left is current selection, together with options to change the status of variables in the selection:



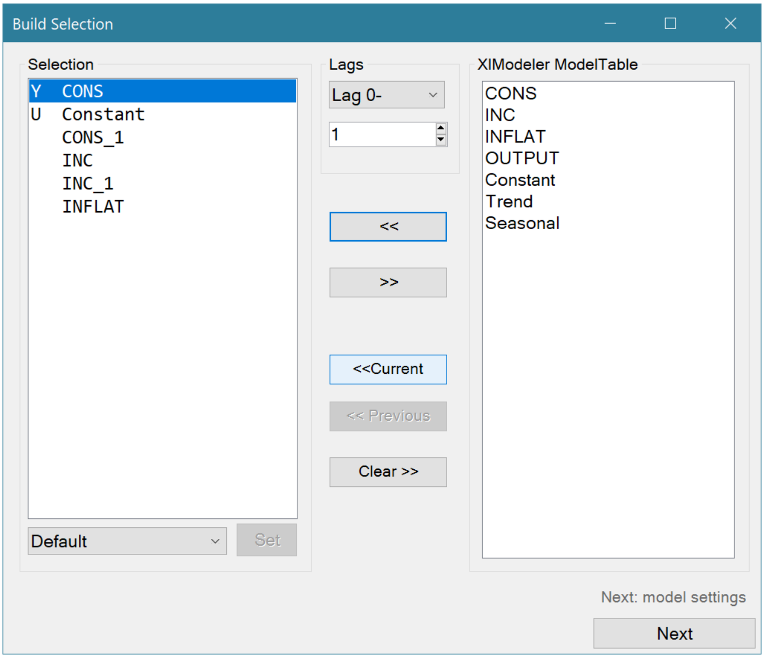
The following actions can be taken in this dialog:

- Select variables in the ModelTabele
- Press >> to add selected variables to the model.
- Double click to add a ModelTable variable directly to the model.
- Remove a variable from the selection by pressing >> or double clicking.
- Empty the *entire* selection by pressing Clear>>.
- On the left-hand side, below the selection (the model formulation), is a drop-down box to change the status of selected variables. It becomes active when a selection variable is selected. For regression model there are three types:
  - **Y** the endogenous (dependent) variable; when starting from an empty model the first is made endogenous by default (and a Constant is added too),
  - **Z** regressor, the default for all other variables,
  - **A** an additional instrument (for instrumental variables estimation, considered later),
  - **U** marks a variable as unrestricted, which will always force it in the model when Autometrics is used. To change status, select one or more variables, then a status type, and click on Set.

The first model to formulate is CONS on a Constant, CONS lagged, INC, INC lagged and INFLAT, as shown below. There are various ways of formulating such a model, including:

- Assuming that the lag selection is set from lag 0 to 1, double click on CONS, INC, INFLAT respectively. Then select INFLAT_1 in the Model list box, and delete (double click, or press the >> button).

- After adding CONS and INC with lags 0 and 1, set the lag length to lag 0 to 0, and add INFLAT.
  This is the model we wish to build:



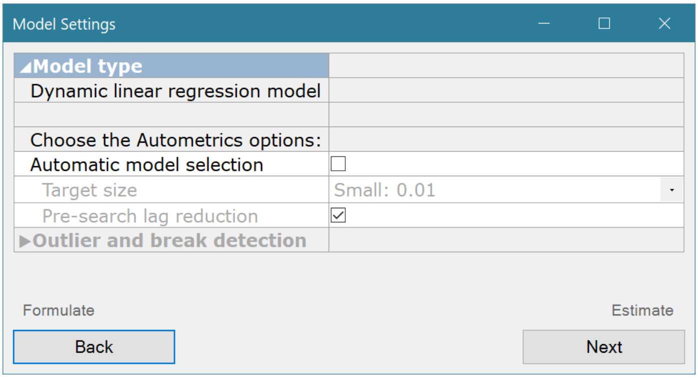
There are three ways to use the lag settings:

- None to add without lags;
- Lag to add only the specified lag;
- Lag 0 to in order to set a lag range.

Note that a Constant is automatically added, but can be deleted if the scale of the variables lets a regression through the origin have meaning. Neither the Constant nor the Trend will be offered for lagging (lagging these would create redundant variables). Seasonals are not used here, but you could add them and delete them if you wish. In that case, you'll see that XlModeler automatically adds the correct number of seasonals (three here as the data are quarterly). It takes the constant term into account; without the constant, four seasonals would have been added. Seasonal is always unity in the first period (first quarter in this case). So Seasonal_1 is one in the second quarter. Remove the seasonals if you've just added them.

2.3 Build a model, step 2: settings

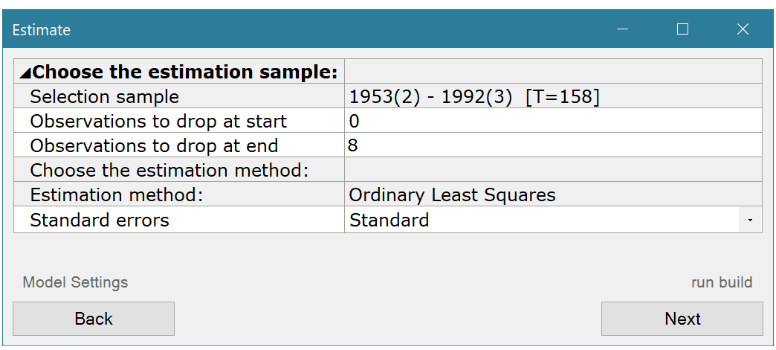
When the variables have been selected, press Next to move to the Model Settings dialog, shown on the next page. For regression models this dialog deals mainly with Autometrics settings. Autometrics is introduced in the next chapter, so press Next again.



2.4 Build a model, step 3: estimation

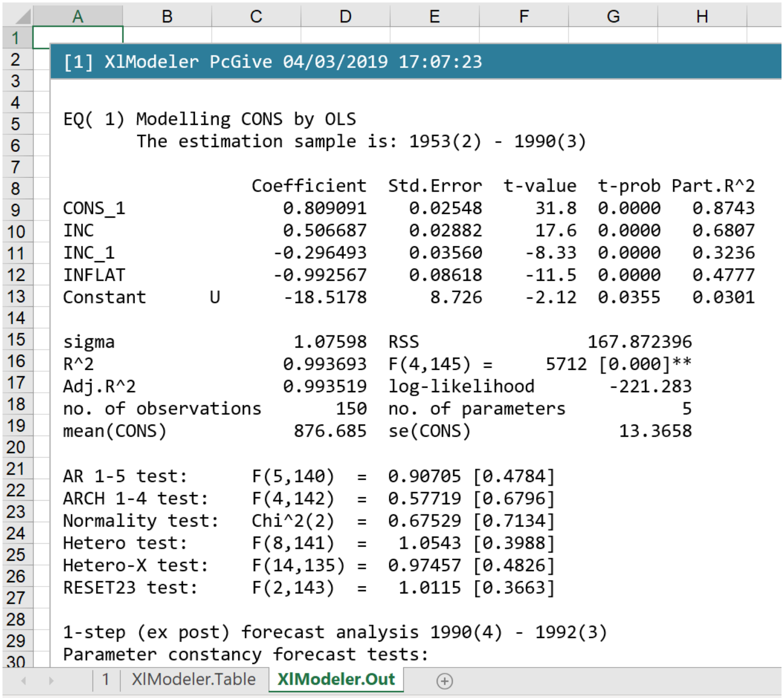
The next step is the the Estimate dialog where the sample period can be set.

The sample period can be adjusted by dropping observations at the start or at the end. Holding back observations at the end allows you to retain some data for static forecasting. Drop *eight* observations at the end:



2.5 Regression model output

Pressing Next again completes the build process by estimating the model. Output appears in a panel in a new sheet, named **XlModeler.Out**:



```
[1] XlModeler PcGive 04/03/2019 17:07:23

EQ( 1) Modelling CONS by OLS
       The estimation sample is: 1953(2) - 1990(3)

                    Coefficient  Std.Error  t-value  t-prob Part.R^2
CONS_1                 0.809091    0.02548     31.8  0.0000   0.8743
INC                    0.506687    0.02882     17.6  0.0000   0.6807
INC_1                 -0.296493    0.03560    -8.33  0.0000   0.3236
INFLAT                -0.992567    0.08618    -11.5  0.0000   0.4777
Constant       U       -18.5178      8.726    -2.12  0.0355   0.0301

sigma                   1.07598  RSS                 167.872396
R^2                    0.993693  F(4,145) =     5712 [0.000]**
Adj.R^2                0.993519  log-likelihood        -221.283
no. of observations         150  no. of parameters            5
mean(CONS)              876.685  se(CONS)               13.3658

AR 1-5 test:      F(5,140)   =  0.90705 [0.4784]
ARCH 1-4 test:    F(4,142)   =  0.57719 [0.6796]
Normality test:   Chi^2(2)   =  0.67529 [0.7134]
Hetero test:      F(8,141)   =   1.0543 [0.3988]
Hetero-X test:    F(14,135)  =  0.97457 [0.4826]
RESET23 test:     F(2,143)   =   1.0115 [0.3663]

1-step (ex post) forecast analysis 1990(4) - 1992(3)
Parameter constancy forecast tests:
```

2.5.1 Equation estimates

The estimated equation has the form:

$$y_t = \mathbf{x}_t' \boldsymbol{\beta} + \epsilon_t, \quad t = 1, \ldots, T,$$

where $\mathbf{x}_t$ contains a '1' for the intercept, y_{t-1} for the lagged dependent variable, as well as the other regressors. Assumptions about the error term are that it has mean 0 and variance which is constant over time:

$$\mathsf{E}[\epsilon_t] = 0, \; \mathsf{V}[\epsilon_t] = \mathsf{E}[\epsilon_t - \mathsf{E}[\epsilon_t]]^2 = \mathsf{E}\left[\epsilon_t^2\right] = \sigma^2.$$

We can write the estimated autoregressive-distributed lag (ADL) model in more detail as:

$$\text{CONS}_t = a_1 \text{CONS}_{t-1} + c + b_0 \text{INC}_t + b_1 \text{INC}_{t-1} + \gamma \text{INFLAT}_t + \epsilon_t. \qquad (2.1)$$

The equation estimation results are written to the **XlModeler.Out** window. The sample period was automatically adjusted for the lags created on CONS and INC. We assume that you have the default options setting, which generates the output as shown. Section **??** discusses further options.

The reported results include coefficient estimates; standard errors; t-values; the squared partial correlation of every regressor with the dependent variable; the squared multiple correlation coefficient (denoted R^2); an F-test on R^2 equalling zero; the equation standard error (σ); the Residual Sum of Squares (RSS). In more detail:

`Coefficient` Thus, the regression coefficients simply show the estimated values of a_1, c, b_0, b_1, γ in (2.1) above. Their interpretation is that a unit increase in INC is associated with a contemporaneous 0.5 unit increase in CONS, but reduced to 0.2 after one more quarter. Importantly, this ignores the lag of CONS, so the impact is really on (CONS-0.8CONS_1).

`Std.Error` Next, the standard errors (SEs) of the coefficients reflect the best estimate of the variability likely to occur in repeated random sampling from the same population: the coefficient $\pm$2SE provides a 95% confidence interval. When that interval does not include zero, the coefficient is often called 'significant' (at the 5% level). The number 2 derives from the assumption that $\hat{\beta}$ has a student-t distribution with $T-k = 150-5 = 145$ degrees of freedom. This, in turn, we know to be quite close to a standard normal distribution, and: $\mathsf{P}\left(|Z| > 2\right) \approx 95\%$ where $Z \sim \mathsf{N}\left(0,1\right)$.

`t-value` The t-value of $\hat{c}$ is the ratio of the estimated coefficient to its standard error:

$$\mathsf{t}_c = \frac{\hat{c}}{\mathsf{SE}\left(\hat{c}\right)},$$

where the latter is obtained from the appripriate diagonal element of the full 5×5 variance matrix of the estimated coefficients. This t-value can be used to test the hypothesis that c is zero (expressed as $\mathsf{H}_0 : c = 0$). Under the current assumptions we reject the hypothesis if $\mathsf{t}_c > 2$ or $\mathsf{t}_c < -2$ (again, using a 95% confidence interval, in other words, a 5% significance level), so values with $|\mathsf{t}| > 2$ are significant. This assumes that the model is statistically well-specified, which we consider below.

`t-prob` The t-probability is the probability of getting a t-value at least as large as the one found, assuming the t-distribution holds. So, an absolute t-value larger than or equal to 2.12 has a probability of 3.55% in a $\mathtt{t}(145)$ distribution.

`Part.R^2` The last statistic in the regression array is the partial r^2. This is the squared correlation between the relevant explanatory variable and the dependent variable (often called regressor and regressand respectively), holding all other variables fixed.

`sigma` The value of $\hat{\sigma}$ is the standard deviation of the residuals, usually called the equation standard error:

$$\hat{\sigma} = \sqrt{\frac{1}{T-k}\sum_{t=1}^{n}\hat{\epsilon}_t^2},$$

for n observations and k estimated parameters (regressors). Since the errors are assumed to be drawn independently from the same distribution with mean zero and constant variance σ, an approximate 95% confidence interval for any one error is $0 \pm 2\hat{\sigma}$. That represents the likely interval from the fitted regression line of the observations. When $\hat{\sigma} = 1.08$, the 95% interval is 4.3% of CONS – the government would not thank you for a value much larger than that, as it knows that consumers' expenditure rarely changes by more than 5% from one year to the next even without your model.

`RSS` RSS is the acronym from residual sum of squares, namely $\sum_{t=1}^{n}\hat{u}_i^2$, which can be useful for hand calculations of tests between different equations for the same variable.

`R^2` R^2 measures the correlation between the actual values CONS_i and the fitted values $\widehat{\text{CONS}}_i$, and is reported immediately below the regression output. When there are several regressors, r^2 and R^2 differ.

`F(.,.)` Moving along the R^2 row of output, the F-test is a test of $\mathsf{R}^2 = 0$. The next item [0.000] is the probability that $\mathsf{F} = 0$, and the ** denotes that the outcome is significant at the 1% level or less.

`Adj.R^2` This is the R^2 adjusted for the number of estimated parameters. Like the R^2, this is not a very useful statistic.

`no. of observations`**...** Finally, the penultimate line gives T and k, while the last line gives the mean and variance of the dependent variable. The variance corresponds to the squared standard deviation:

$$\overline{\text{CONS}} = \frac{1}{n}\sum_{i=1}^{n} \text{CONS}_i,$$

$$[\text{se(CONS)}]^2 = \hat{\sigma}_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(\text{CONS}_i - \overline{\text{CONS}}\right)^2.$$

2.5.2 Test summary: mis-specification tests

Test Summary is reported by default in the output and conducts a summary testing sequence on the residuals for a range of null hypotheses of interest, including: autocorrelation, autoregressive conditional heteroscedasticity (ARCH), the normality of the distribution of the residuals, heteroscedasticity, and functional form mis-specification.

The null hypothesis is in each case the absence of the problem, so it is good to see no significant statistics here. That contributes to our suggestion that the model is statistically well-specified. A further assessment about parameter constancy will be made below. However, if a significant mis-specification is found, it is not necessarily the alternative hypothesis that is the cause of the problem.

Note how easy these tests are to calculate; and to see how informative they are about the match of model and evidence, try computing them when any regressor is dropped (why does dropping INC not lead to rejection?).

2.5.3 Analysis of 1-step forecast statistics

The forecast tests are a Chow test and a forecast $\text{Chi}^2\ (8)$ which is an index of numerical parameter constancy. For H forecasts, values $> 2H$ imply poor *ex ante* accuracy. The third test statistic reported is for the mean of the innovations being zero over the forecast period. This involves the cumulative sum (CUSUM) of the 1-step ahead (recursive) residuals, and has a t-distribution. This information is cut off from the screen capture above, so given here:

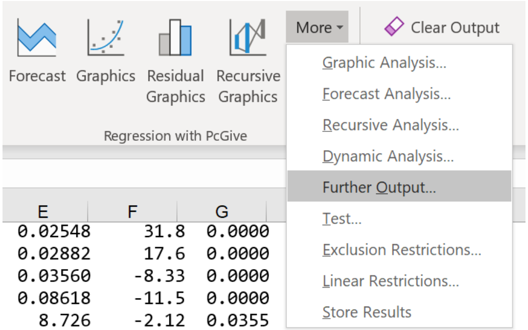
```
1-step (ex post) forecast analysis 1990(4) - 1992(3)
Parameter constancy forecast tests:
Forecast  Chi^2(8) = 9.3241 [0.3157]
```

```
Chow      F(8,145) = 1.1500 [0.3337]
CUSUM     t(7)     =-0.9719 [0.3635]  (zero forecast innovation mean)
```

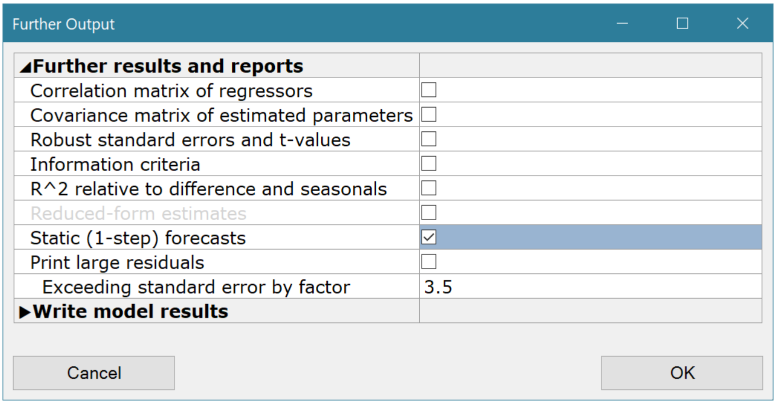
so these test do not detect any non-constancy for the forecast period.

Later, we will graph the outcomes, forecasts and the error bars for ±2 standard errors of the 1-step forecasts.

To see the full results, use More/Further Output from the ribbon



and select Static (1-step) forecasts:



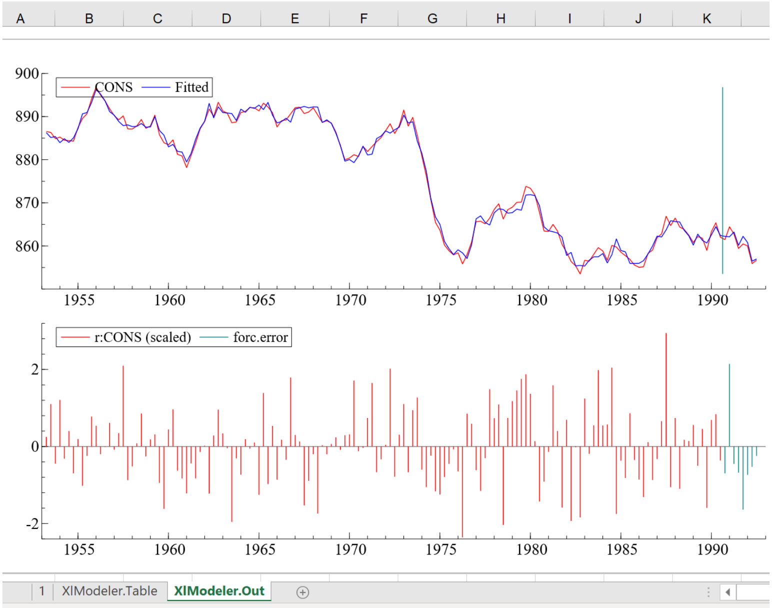

```
1-step forecasts for CONS (SE with parameter uncertainty)
   Horizon      Forecast          SE       Actual        Error    t-value
    1990-4       862.235       1.097      861.484     -0.75153     -0.685
    1991-1       862.136       1.090      864.444       2.3080      2.117
    1991-2       863.237       1.085      862.750     -0.48749     -0.449
    1991-3       860.146       1.091      859.413     -0.73240     -0.671
    1991-4       862.243       1.106      860.480      -1.7626     -1.594
    1992-1       860.796       1.092      860.002     -0.79344     -0.727
    1992-2       856.477       1.113      855.908     -0.56831     -0.511
    1992-3       856.995       1.089      856.731     -0.26431     -0.243
   mean(Error)  =      -0.38151   RMSE =       1.1616
   SD(Error)    =        1.0972   MAPE =      0.11129
```

The RMSE and MAPE are discussed in §2.13.

2.6 Graphics and Residual Graphics

The next major step is to do a graphical evaluation of the estimated model. There are convenient buttons for a quick analysis, and additional options to create advanced graphs.

First using the Graphics button. This creates two panels inside one graph, shown as a screen capture here:



At the top are actual and fitted values, where the last observations are static forecasts and outcomes. At the bottom are the residuals, with the last eight bars the forecast errors.

This graph can be copied into a Word document.

Next is the Residual Graphics button, with the result shown in Figure 2.1.

2.7 Further graphical analysis

For a wider range of graphs, select the Graphic Analysis dialog from the More menu (or press the toolbar button). Mark the first six items:

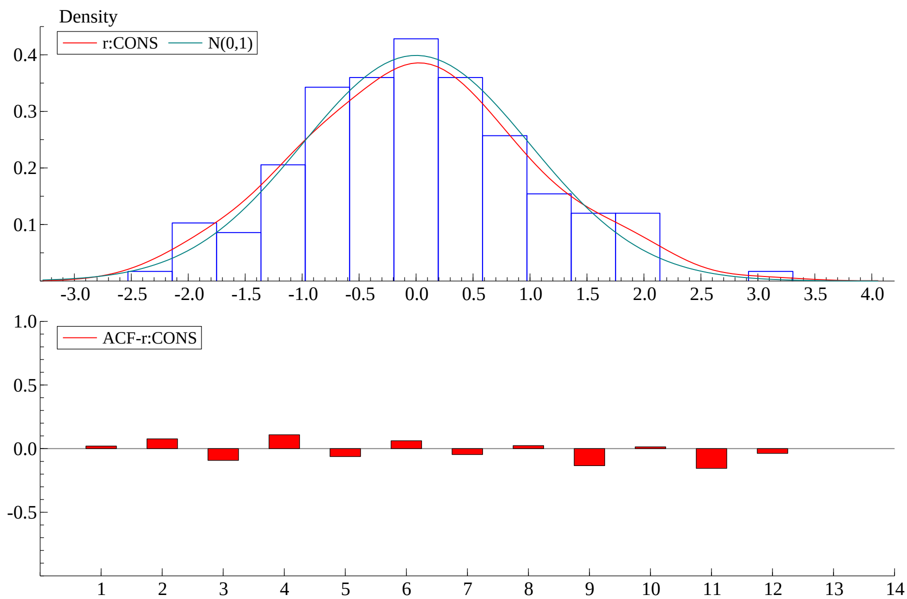
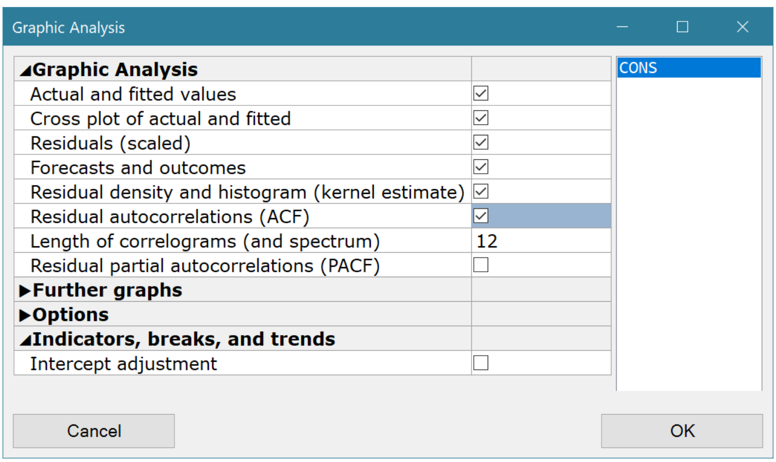


Figure 2.1 Residual Graphics of CONS model



The dialog lets you plot or cross-plot the actual and fitted values for the whole sample, the residuals scaled by σ, so that values outside the range $[-2, +2]$ suggest outlier problems, the forecasts, and some graphical diagnostic information about the residuals (their spectrum, correlogram, histogram, density and cumulative distribution). The forecast period start is marked by a vertical line (see Figure 2.2). Notice the good fit: the earlier high R^2, and good Chow test are consistent with this. As before, any graphs can be saved for later recall, editing and printing.

Accept the dialog, and the graphs appear in the XlModeler Graphics window, as in Figure 2.2. There are two new graphs. The first is the correlogram, which extends the idea behind the DW test to plot the correlations between successive lagged residuals

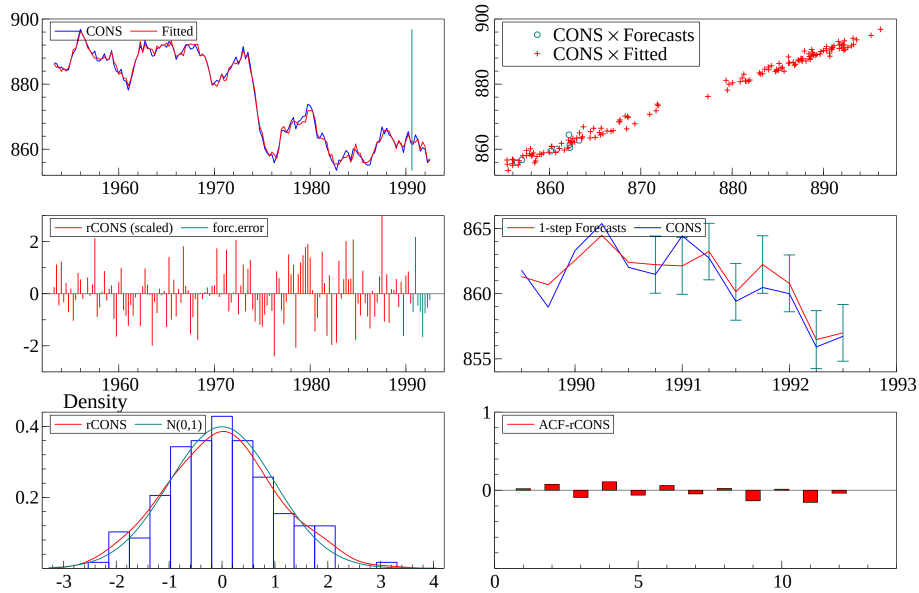


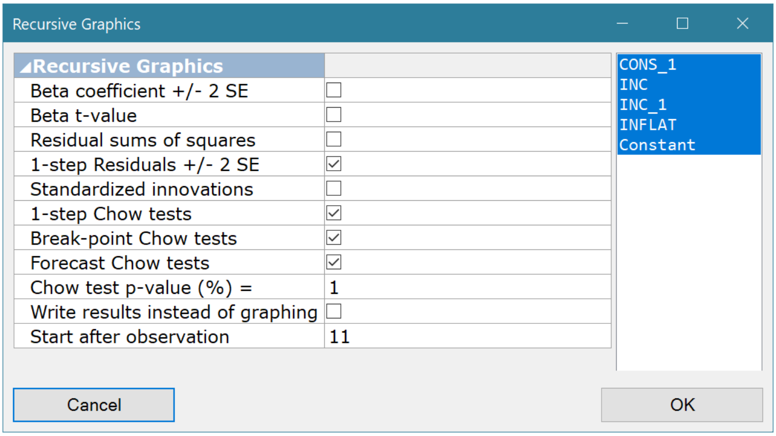
Figure 2.2 Graphical evaluation of CONS model

(that is, the correlation of $\hat{\epsilon}_t$ with $\hat{\epsilon}_{t-1}$, then with $\hat{\epsilon}_{t-2}$, $\hat{\epsilon}_{t-3}$ and so on up to $\hat{\epsilon}_{t-12}$). A random (independent) residual would have most such correlations close to zero: visually, the dependence between successive residuals is small. The second plots the forecasts which we printed earlier, with the error bands changed to error fans.

2.8 Recursive estimation

We had already noted that the model appeared constant over the forecast period. Our next topic is recursive estimation: the logic is simply to repeatedly drop the last observation and re-estimate to see if the results remain as expected. So the sample shrinks $T, T-1, T-2, ...$ until it gets too small for meaningful results. The main output will be graphs of coefficients, $\hat{\sigma}$ etc. over the changing sample size. This is a powerful way to study parameter constancy (especially in its absence!). We sometimes refer to recursively applied OLS as RLS.

Quick Recursive Graphics dialog is activated from the toolbar button. A more detailed analysis is available from the Recursive Analysis menu):



The right column has all the variables to be plotted in beta coefficients and beta t-values. Select the statistics which are to be plotted: beta coefficients, beta t-values, 1-step residuals, and all three Chow tests. To reduce the number of graphs, we have excluded the coefficient on `INFLAT`.

First, the graph of the coefficient of CONS_1 over the sample in Figure 2.3 shows that after 1978, $\hat{\beta}_t$ lies almost outside of the previous confidence interval which an investigator pre-1974 would have calculated as the basis for forecasting. Other coefficients are also non-constant. Further, the 1-step residuals show one outlier around 1987.

The 1-step residuals are

$$\tilde{u}_t = y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}_t$$

and they are plotted with $\pm 2\hat{\sigma}_t$ shown on either side of zero. Thus $\tilde{u}_t$ which are outside of the error bars are either outliers or are associated with changes in $\hat{\sigma}$. The full sample residuals are:

$$\hat{u}_t = y_t - \mathbf{x}_t'\hat{\boldsymbol{\beta}}_T$$

where $\hat{\boldsymbol{\beta}}_T$ is the full-sample OLS estimate. Graphic analysis plots these, scaled by the full sample $\hat{\sigma}_T$.

Further summary graphs are the Chow tests, which are all scaled by their 1% critical value (which becomes the line at unity). The 1-step Chow tests evaluate the one step ahead forecasts, whereas the forecast (or Nup) tests evaluate the forecasts at each point relative to the estimates at the start of the recursive plots. So the forecast Chow tests have an expanding forecast horizon. Finally, in the Ndown, or breakpoint, Chow test each point is the value of the Chow F-test for that date against the final period, here 1990(3), again scaled by its 1% critical value, so the forecast horizon N is decreasing from left to right (hence the name Ndn tests). Figure 2.3 illustrates: the critical value can be set at any desired probability level. The breakpoint Chow test shows a failure around 1974 (the oil crisis ...).

Peruse other options as you wish: see how the standardized innovations often highlight the outliers, or how the residual sums of squares confirm that a break occurred

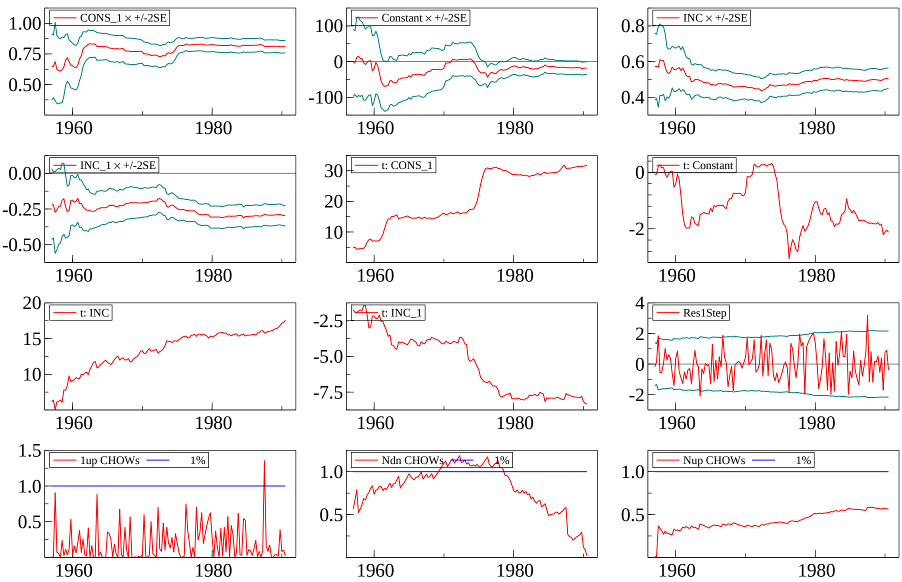


Figure 2.3 Recursive least squares graphical constancy statistics

in 1974. Any of these graphs can be edited and printed, or saved for later recall and printing.

Next, select the Graphic analysis dialog and look at the full sample residuals:

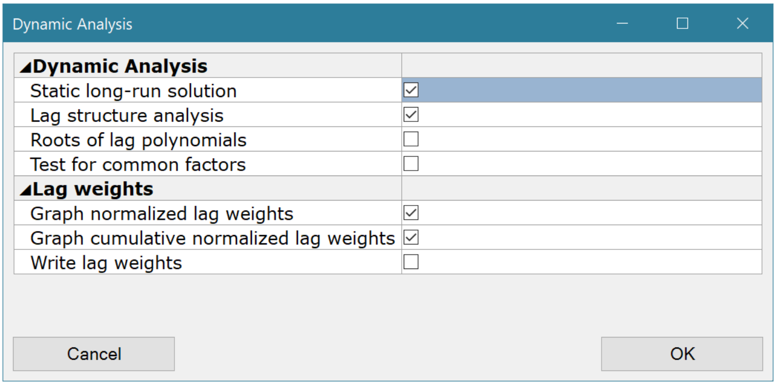
$$\hat{u}_t = y_t - \mathbf{x}_t' \hat{\boldsymbol{\beta}}_T$$

where $\hat{\boldsymbol{\beta}}_T$ is the usual full-sample OLS estimate. The full-sample estimates somewhat smooth the outliers evident in the recursive figures, so now the largest is not much more than 3.5 standard errors (partly because $\hat{\sigma}$ increased by about 50% over the sample).

Note that recursive estimation can also be used to check the constancy of pre-tests such as (Augmented) Dickey–Fuller tests for the order of integration of a time series, and the recursive graphs may help to discriminate between genuine unit roots and autoregressive coefficients driven towards unity by a failure to model a regime shift.

2.9 Dynamic analysis

Next, activate Dynamic Analysis from the Test menu. Select Static long-run solution, Lag structure analysis, and both Graph normalized weights and Graph cumulative normalized weights, as shown:



The dynamic analysis commences with the long-run solution. The solved long-run model (or static solution) is calculated, together with the relevant standard errors as follows. Write the dynamic equation as

$$a\,(L)\,y_t = b\,(L)\,x_t + \epsilon_t,$$

where L is the lag operator so that $Lx_t = x_{t-1}$ and $b(L) = \sum_{i=0}^{n} b_i L^i$ is a scalar polynomial in L of order n, the longest lag length. Similarly, $a(L) = \sum_{i=0}^{n} a_i L^i$, with $a_0 = -1$. With $a(1) = \sum_{i=0}^{n} a_i$ (that is, $a(L)$ evaluated at $L = 1$), then if $a(1) \neq 0$ the long run is:

$$y = \frac{b\,(1)}{a\,(1)} x = Kx.$$

Under stationarity (or cointegration inducing a stationary linear relation), standard errors for derived coefficients like K can be calculated from those of $a\,(\cdot)$ and $b\,(\cdot)$. Here the long-run coefficients are well determined, and the null that they are all zero (excluding the constant term) is rejected.

```
Solved static long run equation for CONS
                    Coefficient  Std.Error   t-value   t-prob
INC                     1.10102    0.04534      24.3    0.000
INFLAT                 -5.19917     0.5558     -9.35    0.000
Constant               -96.9979      40.68     -2.38    0.018
Long-run sigma = 5.63611

ECM = CONS + 96.9979 - 1.10102*INC + 5.19917*INFLAT;
WALD test: Chi^2(2) = 824.782 [0.0000] **
```

Next, the lag polynomials are analyzed, listing the individual coefficients a_0, a_1, etc. (normalized so that $a_0 = -1$), followed by their sum $a(1)$, $b(1)$, etc. and their standard errors (remember that the standard error of the sum is not simply the sum of the standard errors!):

```
Analysis of lag structure, coefficients:
              Lag 0     Lag 1       Sum   SE(Sum)
CONS             -1     0.809    -0.191    0.0255
Constant      -18.5         0     -18.5      8.73
INC           0.507    -0.296      0.21    0.0313
INFLAT       -0.993         0    -0.993    0.0862
```

This is followed by the F-tests of the joint significance of each variable's lag polynomial:

```
Tests on the significance of each variable
Variable      F-test          Value [  Prob]     Unit-root t-test
CONS          F(1,145) =     1008.5 [0.0000]**       -7.4931**
Constant      F(1,145) =      4.503 [0.0355]*
INC           F(2,145) =     155.67 [0.0000]**        6.7226
INFLAT        F(1,145) =     132.64 [0.0000]**       -11.517

Tests on the significance of each lag
Lag 1         F(2,145) =     617.11 [0.0000]**
```

The hypothesis that $a(1) = 0$ can be rejected, with a XlModeler unit-root test value of -7.49 (or $-0.191/0.0255$ from the previous output). The two stars mark significance, suggesting cointegration between the variables in the model in levels (see Banerjee, Dolado, Galbraith, and Hendry, 1993, or Johansen, 1995). Finally, tests on the significance of each lag length are provided (here we deleted three columns with zeros): The unit-root t-test (also called the XlModeler unit-root test) does not in fact have a t-distribution, but the marked significance (* for 5%, ** for 1%, dependent variable only) is based on the correct critical values, see Banerjee, Dolado, and Mestre (1998).

Since we also chose Lag weights, there are four new graphs in the XlModeler output, as in Figure 2.4.

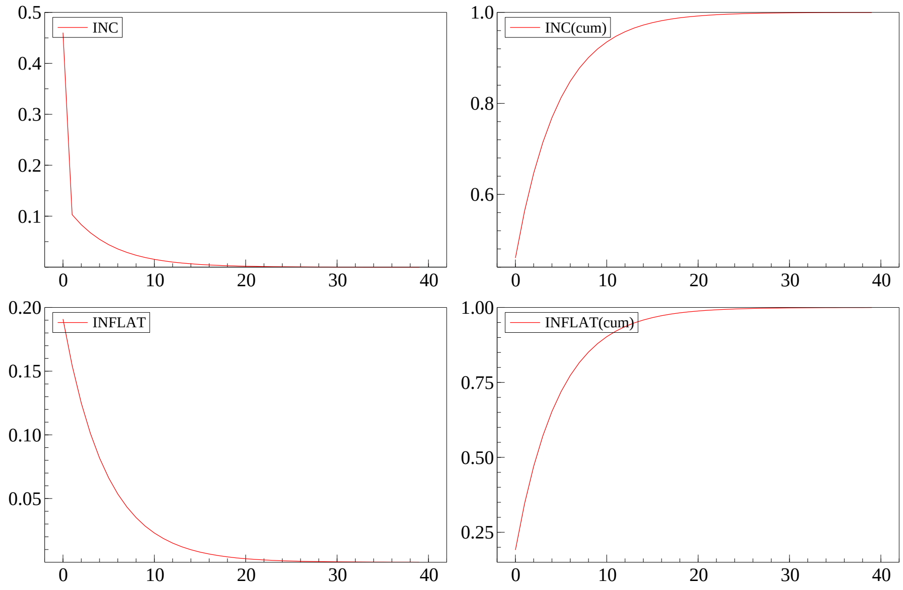
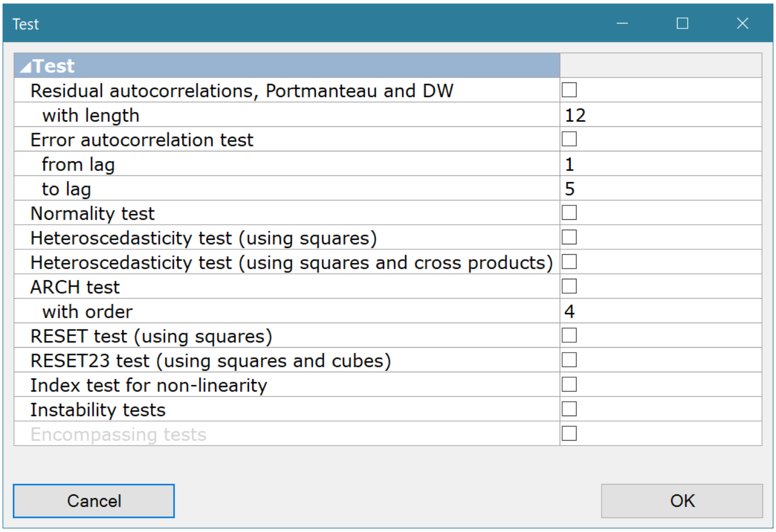


Figure 2.4 Lag weights from CONS model

2.10 Mis-specification tests

The test summary was already considered above. Tests can also be undertaken individually, or in different groups from that embodied in the test summary. From the Test menu, select Test, which brings up the Test dialog:



Any or all available tests can be selected.

Tip The default values for the lag length of the AR and ARCH tests are based on the data frequency and the sample size. Different lag lengths can be selected in the Test dialog.

The output is rather more extensive than with summary tests. For example, the error autocorrelation test (or AR test) and ARCH test produce:

```
Error autocorrelation coefficients in auxiliary regression:
  Lag Coefficient  Std.Error
    1     0.054381     0.0909
    2     0.071683    0.08987
    3    -0.088316    0.08768
    4      0.12258    0.08856
    5    -0.051648     0.0885
RSS = 162.605  sigma = 1.16146

Testing for error autocorrelation from lags 1 to 5
 Chi^2(5) = 4.7067 [0.4527]  and F-form F(5,140) = 0.90705 [0.4784]

ARCH coefficients:
  Lag Coefficient  Std.Error
    1    -0.065174    0.08399
    2    -0.077844    0.08388
    3     0.063463    0.08394
    4     0.034848    0.08416
RSS = 345.559  sigma = 1.55997
```

```
Testing for error ARCH from lags 1 to 4
ARCH 1-4 test:    F(4,142)  =  0.57719 [0.6796]
```

The DW-statistic (Durbin–Watson) is not printed by default. Instead, it is available as the first entry on the Test dialog. But note that the assumptions needed to justify the application of the DW test in economics are rarely satisfied.

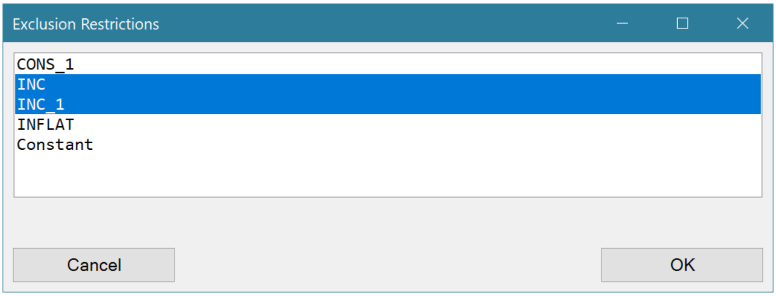
Similarly, the normality test leads to the low-order moments being reported. The density of the scaled residuals was shown in Figure 2.2 and revealed slight skewness and somewhat fatter tails than the standard normal distribution. These mis-specification test outcomes are satisfactory, consistent with the equation being a congruent model, so we now consider some specification tests.

Note that you can use Test/Store to store residuals and fitted values from the regression in the workbook.

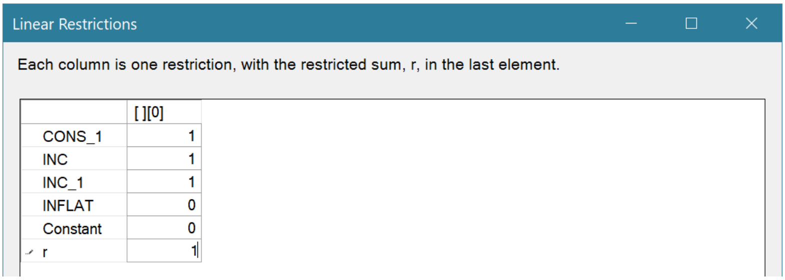
2.11 Specification tests

2.11.1 Exclusion, linear and general restrictions

First, we test whether a subset of the coefficients is zero. Choose Exclusion Restrictions from the Test menu to test whether one or more coefficients are zero. At the dialog mark INC and INC_1 and accept:



Before looking at the subset test result, we shall also do a linear restrictions test on homogeneity of CONS with respect to INC. This time, select Linear Restrictions. To complete, edit the restrictions as follows:



This formulates one restriction. The last element is the r vector, specifying what the restriction should add up to. In terms of (2.1) the restrictions are:

$$(1\ 0\ 1\ 1\ 0)\begin{pmatrix} a_1 \\ c \\ b_0 \\ b_1 \\ \gamma \end{pmatrix} = a_1 + b_0 + b_1 = 1.$$

The results of the two tests are:

```
Test for excluding:
[0] = INC
[1] = INC_1
Subset F(2,145) =   155.67 [0.0000]**

Test for linear restrictions (Rb=r):
R matrix
       CONS_1          INC         INC_1        INFLAT     Constant
       1.0000       1.0000        1.0000       0.00000      0.00000
r vector
       1.0000
LinRes F(1,145) =   4.0348 [0.0464]*
```

The output of the homogeneity test shows slight evidence of rejection of long-run homogeneity if conventional critical values are used. The previous exclusion test reveals strong rejection of the null.

2.11.2 Test for common factors

Testing for common factors (COMFAC; is part of Dynamic analysis. It is also a specification test, motivating its inclusion here. When the dynamic equation is:

$$a\,(L)\,y_t = b\,(L)\,x_t + c\,(L)\,z_t + \epsilon_t,$$

COMFAC involves testing whether $a(L) = a(1 - \rho L)$ when $b(L) = b(1 - \rho L)$ and $c(L) = c(1 - \rho L)$ so that $(1 - \rho L)$ is the factor of the lag polynomials in common. COMFAC is discussed by Hendry and Mizon (1978).

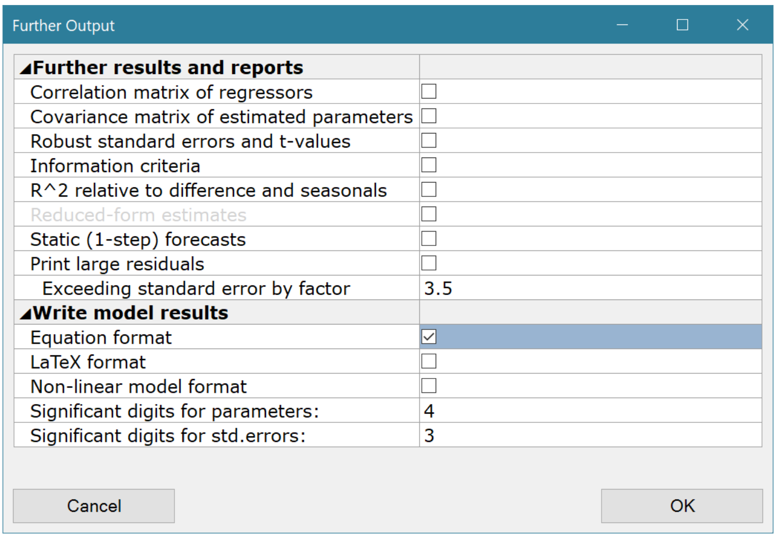
To select COMFAC tests, the minimum lag length must be unity for all non-redundant variables (variables that are redundant when lagged can occur without lags: XlModeler notices the Constant and Trend if such terms occur). First, we must revise the model to have one lag on INFLAT: return to build a Regression Model dialog, mark INFLAT in the database, and add it to the model using a lag length of one. XlModeler notices that current INFLAT is already in the model and doesn't add it a second time. Estimate over the previously selected sample (holding back 8 observations at the end). Now select Dynamic analysis, and mark Test for common factors. Since the lag polynomials are first-order, only the Wald test of one common-factor restriction is presented following the roots of the lag polynomials. Here the restriction is rejected so the dynamics do not have an autoregressive error representation, matching the very different roots of the lag polynomials. The output is:

```
COMFAC Wald test table, COMFAC F(2,144) = 51.3999 [0.0000] **
Order   Cumulative tests                Incremental tests
  1     Chi^2(2) =  102.8 [0.0000]**    Chi^2(2) =  102.8 [0.0000]**
```

The remainder of this chapter uses the model without lagged INFLAT, so re-estimate the previous model (still with 8 static forecasts).

2.12 Further Output

Other output formats may prove more convenient for direct inclusion in final reports. To make XlModeler write the output in equation format, for example, activate the Further Output, (on the Test menu) and mark as shown:



Reporting only the equation format, this produces:

```
CONS =   + 0.8091*CONS_1 + 0.5067*INC - 0.2965*INC_1 - 0.9926*INFLAT
(SE)      (0.0255)         (0.0288)     (0.0356)       (0.0862)
         - 18.52
          (8.73)
```

2.13 Forecasting

To end this chapter, we briefly compare dynamic and static forecasts for this model. When we started, 8 observations were kept for static (1-step) forecasts, and these were listed in §2.5.3 and graphed in Fig. 2.2d.

For a quick forecast press the `Forecast` button. For quarterly data, this will produce 3 years of dynamic (out-of-sample) forecasts. Because the model is conditional on INC and INFLAT, we cannot forecast any further when these run out. The forecasts are graphed as well as written in detail in the **XlModeler.Out** sheet.

```
Dynamic (ex ante) forecasts for CONS (SE based on error variance only)
 Horizon   Forecast      SE    Actual      Error    t-value     -2SE      +2SE
 1990(4)    862.235   1.076   861.484   -0.75153     -0.698    860.08    864.39
 1991(1)    862.744   1.384   864.444     1.7000      1.228    859.98    865.51
 1991(2)    861.862   1.553   862.750    0.88793      0.572    858.76    864.97
 1991(3)    859.427   1.654   859.413  -0.013983     -0.008    856.12    862.74
 1991(4)    862.254   1.717   860.480    -1.7739     -1.033    858.82    865.69
 1992(1)    862.231   1.757   860.002    -2.2287     -1.268    858.72    865.75
 1992(2)    858.280   1.783   855.908    -2.3715     -1.330    854.71    861.85
 1992(3)    858.914   1.800   856.731    -2.1831     -1.213    855.31    862.51
 1992(4)       .NaN   1.811      .NaN       .NaN       .NaN      .NaN      .NaN
 1993(1)       .NaN   1.818      .NaN       .NaN       .NaN      .NaN      .NaN
 1993(2)       .NaN   1.822      .NaN       .NaN       .NaN      .NaN      .NaN
 1993(3)       .NaN   1.825      .NaN       .NaN       .NaN      .NaN      .NaN
 mean(Error) =      -0.84186    RMSE =       1.6862
 SD(Error)   =        1.4611    MAPE =      0.17320
```

Two summary statistics are reported in addition to the mean and standard deviation of the error (over those that are available). The first is the Root Mean Square Error:

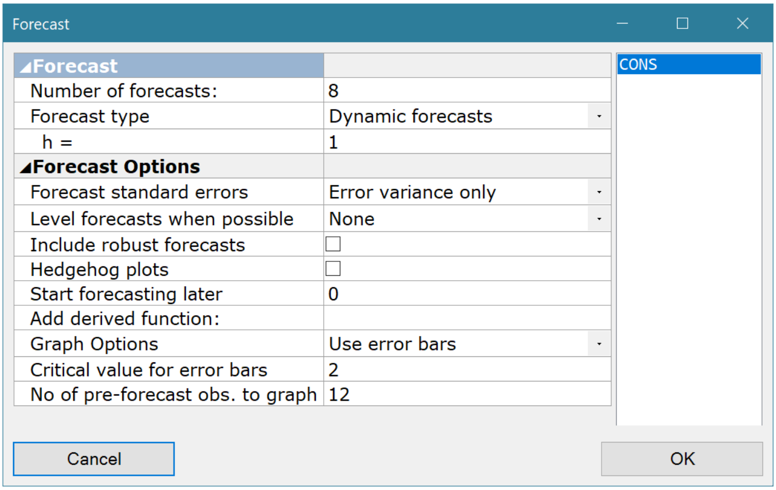
$$\mathrm{RMSE} = \left[\frac{1}{H}\sum_{t=1}^{H}(y_t - f_t)^2\right]^{1/2},$$

where the forecast horizon is H (8 here), y_t the actual values, and f_t the forecasts. The second statistic is the Mean Absolute Percentage Error:

$$\mathrm{MAPE} = \frac{100}{H}\sum_{t=1}^{H}\left|\frac{y_t - f_t}{y_t}\right|.$$

Both are measures of forecast accuracy, see, e.g. Makridakis, Wheelwright, and Hyndman (1998, Ch. 2). Note that the MAPE can be infinity if any $y_t = 0$, and is different when the model is reformulated in differences. For more information see Clements and Hendry (1998).

More options are available through Forecast Analysis:



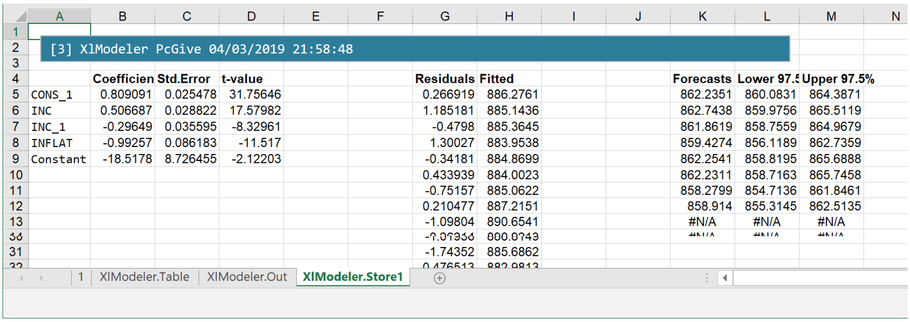
The one-step forecasts are also called *ex-post* forecasts: they require actual data of all explanatory variables. To obtain the one-step forecast of CONS for 1990Q4, we need to know the INC and INFLAT values for 1990Q4, and CONS from the previous period. The next static forecast is again based on observed values for INC, INFLAT and previous CONS. For pure forecasting purposes, we need to make *dynamic forecasts*, usually requiring forecasts of all explanatory variables as well. One solution is to switch to multiple-equation dynamic modelling, and to make INC and INFLAT endogenous in a system such that all the variables are jointly forecast. That is beyond our current scope, but we can do something comparable by at least using forecasted values of CONS when available. In a simple autoregressive model $y_t = \beta y_{t-1} + \epsilon_t$, writing $\hat{y}_t$ for forecasted values, and assuming that $T+1$ is the first forecast period:

Forecast horizon	Static Forecast	Dynamic Forecast
$T+1$	$\hat{y}_{T+1} = \hat{\beta} y_T$	$\hat{y}_{T+1} = \hat{\beta} y_T$
$T+2$	$\hat{y}_{T+2} = \hat{\beta} y_{T+1}$	$\hat{y}_{T+2} = \hat{\beta} \hat{y}_{T+1}$
$T+3$	$\hat{y}_{T+3} = \hat{\beta} y_{T+2}$	$\hat{y}_{T+3} = \hat{\beta} \hat{y}_{T+2}$

The first forecast is the same, but thereafter the forecasts differ.

2.14 Store Results

Note that you can use Store Results to store residuals, fitted values, and dynamic forecasts from the regression in the Excel workbook. This creates a new sheet **XlModeler.Store1** (and then Store2, Store3, ...):



[3] XlModeler PcGive 04/03/2019 21:58:48

	Coefficien	Std.Error	t-value	Residuals	Fitted	Forecasts	Lower 97.5	Upper 97.5%
CONS_1	0.809091	0.025478	31.75646	0.266919	886.2761	862.2351	860.0831	864.3871
INC	0.506687	0.028822	17.57982	1.185181	885.1436	862.7438	859.9756	865.5119
INC_1	-0.29649	0.035595	-8.32961	-0.4798	885.3645	861.8619	858.7559	864.9679
INFLAT	-0.99257	0.086183	-11.517	1.30027	883.9538	859.4274	856.1189	862.7359
Constant	-18.5178	8.726455	-2.12203	-0.34181	884.8699	862.2541	858.8195	865.6888
				0.433939	884.0023	862.2311	858.7163	865.7458
				-0.75157	885.0622	858.2799	854.7136	861.8461
				0.210477	887.2151	858.914	855.3145	862.5135
				-1.09804	890.6541	#N/A	#N/A	#N/A
				-1.74352	885.6862			

All the results of this chapter have been saved in `data-1.xlsx`, which is stored in your user folder under XlModeler.

In the next chapter we consider strategies for model reduction and the automatic facilities that XlModeler offers to the applied modeller. This is a major time-saving device.

Chapter 3

Regression Model using Automatrics

3.1 Introduction

We now turn to what is perhaps the most useful part of regression models using XlModeler and PcGive: automatic model selection with *Autometrics*. The objective is to let the computer do a large part of what was done by hand in the previous chapter. XlModeler will be able to find a model much quicker than we can. Of course, there is always the option to do the model selection by hand — but it will be quite a challenge to beat *Autometrics*.

Autometrics is a computer implementation of general-to-specific modelling, see Doornik (2009), Doornik (2008). This follows on from Hendry and Krolzig (1999) and Hoover and Perez (1999). There is now considerable Monte Carlo simulation evidence that *Gets* performs well, selecting a model from an initial general specification almost as often as the same criteria would when applied to the DGP (since test size leads to false rejections, and non-unit power to false acceptances of the null even when the analysis commences from the 'truth'). A separate book, Hendry and Doornik (2014), treats *Autometrics* and econometric model selection in general, including saturation-based estimators such as impulse indicator saturation (IIS) and step indicator saturation (SIS).

3.2 The problems of simple-to-general modelling

While the models of the previous chapter were mainly selected as illustrations of how to use XlModeler, they highlighted four important issues:

1. Powerful tests can reveal model inadequacies: it is not sensible to skip testing in the hope that the model is valid.
2. A reject outcome on any test invalidates all earlier inferences, rendering useless the time spent up to then – empirical research becomes highly inefficient if done that way.
3. Once a problem is revealed by a test, how do you proceed? It is a dangerous *non sequitur* to adopt the alternative hypothesis of the test which rejected: will you be tempted to do this with residual autocorrelation, by assuming it is error autoregression?
4. What can be done if two or more statistics reject? Which has caused what? Do both or only one need to be corrected? Or should third factors be sought?

As discussed by Hendry and Doornik (2014) and many others, the whole paradigm of postulating a simple model and seeking to generalize it by searching for significant variables (as embodied by stepwise regression) or test rejections is suspect, and in fact makes sub-optimal use of XlModeler's structure and functioning. Let us now switch to its mode of general-to-specific modelling.

3.3 Formulating general models

We continue with `data.xlsx` in this chapter, starting with a clean modelling sheet. So reopen the original `data.xlsx` if you want your model numbering to coincide with the output presented in this chapter.

Turn to build a Regression Model to create a general specification. In substantive research, the starting point should be based on previous empirical research evidence (to test in due course that earlier findings are parsimoniously encompassed), economic (or other relevant subject matter) theory, institutional knowledge, the data frequency – and common sense. Here, we base the initial model on Davidson, Hendry, Srba, and Yeo (1978) (denoted DHSY below) and begin by formulating an equation with CONS, INC, INFLAT and Constant as its basic variables (you could add in OUTPUT too if you like, but logic suggests it should be irrelevant given income).

Choose CONS, INC, and INFLAT with two lags each: please note that we are still only illustrating – in practice, five lags would be a better initial lag length for quarterly data, which we will do later using Autometrics. Do not retain any forecasts for this run and select full sample OLS:

```
EQ( 1) Modelling CONS by OLS
       The estimation sample is: 1953(3) - 1992(3)

                Coefficient  Std.Error  t-value  t-prob Part.R^2
CONS_1             0.823338    0.08223     10.0   0.000   0.4038
CONS_2           -0.0315462    0.07151   -0.441   0.660   0.0013
INC                0.500117    0.02922     17.1   0.000   0.6643
INC_1             -0.295872    0.05568    -5.31   0.000   0.1602
INC_2             0.0255575    0.04471    0.572   0.568   0.0022
INFLAT            -0.844115     0.2521    -3.35   0.001   0.0704
INFLAT_1         -0.0801516     0.4348   -0.184   0.854   0.0002
```

```
INFLAT_2          -0.137750    0.2633   -0.523   0.602   0.0018
Constant    U      -20.7434     9.070    -2.29   0.024   0.0341

sigma               1.09027  RSS               175.92662
R^2                0.993857  F(8,148) =     2993 [0.000]**
Adj.R^2            0.993524  log-likelihood     -231.708
no. of observations     157  no. of parameters         9
mean(CONS)           875.78  se(CONS)             13.5487
```

Scan the output, noting the coefficient estimates *en route* (e.g. four t-values are small). Select the dynamic analysis to compute the static long-run solution:

```
Solved static long run equation for CONS
                  Coefficient  Std.Error  t-value  t-prob
INC                   1.10372    0.04237     26.0   0.000
INFLAT               -5.10074     0.5484    -9.30   0.000
Constant             -99.6282      38.05    -2.62   0.010
Long-run sigma = 5.23645

ECM = CONS - 1.10372*INC + 5.10074*INFLAT + 99.6282;
WALD test: Chi^2(2) = 981.088 [0.0000] **
```

Note the coefficient values (for example, INC is close to unity, INFLAT to -5) and their small standard errors (so INC is apparently significantly different from unity).

3.4 Analyzing general models

The analysis of the lag structure is now more interesting: the unit-root t-tests show that the three basic variables matter as long-run levels (less so if very long lags were selected initially), which rejects a lack of cointegration. The F-tests on the (whole) lag polynomials show that each also matters dynamically. However, lag length 2 is irrelevant, whereas the first lag cannot be removed without a significant deterioration in fit.

```
Analysis of lag structure, coefficients:
            Lag 0    Lag 1    Lag 2      Sum  SE(Sum)
CONS           -1    0.823  -0.0315   -0.208   0.0322
Constant    -20.7        0        0    -20.7     9.07
INC           0.5   -0.296   0.0256     0.23    0.038
INFLAT     -0.844  -0.0802   -0.138    -1.06    0.132

Tests on the significance of each variable
Variable     F-test        Value [  Prob]     Unit-root t-test
CONS         F(2,148) =   306.93 [0.0000]**       -6.4716**
INC          F(3,148) =   102.74 [0.0000]**        6.0458
INFLAT       F(3,148) =   32.254 [0.0000]**        -8.039
Constant     F(1,148) =   5.2302 [0.0236]*

Tests on the significance of each lag
Lag 2        F(3,148) =  0.17158 [0.9155]
Lag 1        F(3,148) =   38.745 [0.0000]**
```

```
Tests on the significance of all lags up to 2
Lag 2 - 2    F(3,148) =  0.17158 [0.9155]
Lag 1 - 2    F(6,148) =   207.76 [0.0000]**
```

These four perspectives on the model highlight which reductions are consistent with the data, although they do not tell you in what order to simplify. That issue can be resolved in part by more experienced researchers (for one example see the discussion in Hendry, 1987). For the moment, we will follow a sequential simplification route, although generally it is better to transform to near orthogonality prior to simplification.

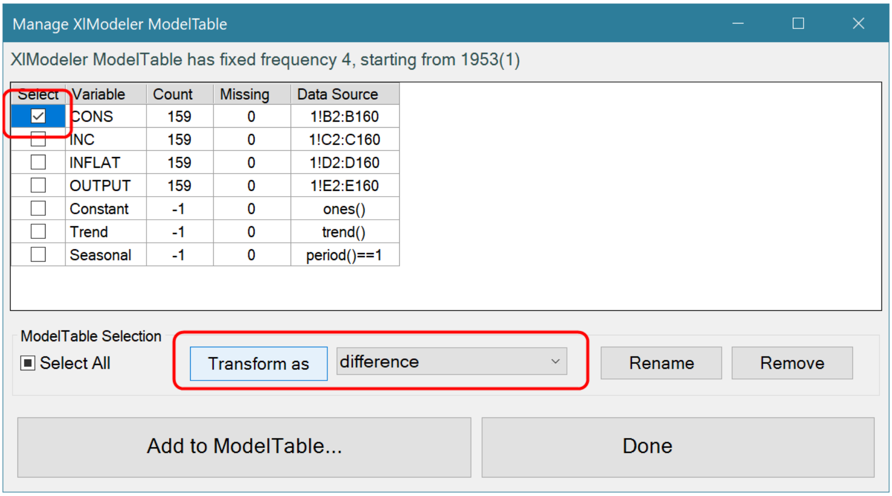
Can we trust the tests just viewed? The natural attack on that issue is to test all of the congruency requirements listed in the Help: so test using the test summary. Many of these tests will already have been conducted during earlier tutorials. The residual plot looks normal, and no test rejects, although either of the autocorrelation or RESET tests suggests a possible problem may be lurking in the background (the former option gives significant negative autocorrelation possibly owing to overfitting – keep an eye on how that evolves as simplification proceeds). COMFAC accepts that one common factor can be extracted (matching the insignificant 2^{nd} order lag, which would imply that the common factor had a coefficient of zero) but strongly rejects extracting two. The omitted variables[1] test reveals that OUTPUT is indeed irrelevant. And the linear restrictions test confirms that long-run homogeneity of CONS with respect to INC is rejected at the 5% level. Tentatively, therefore, we accept the general or statistical model as data-congruent, with no need for the second lag.

```
AR 1-5 test:       F(5,143)  =   2.1861 [0.0589]
ARCH 1-4 test:     F(4,149)  =   1.0123 [0.4030]
Normality test:    Chi^2(2)  =   1.6495 [0.4384]
Hetero test:       F(16,140) =  0.74629 [0.7425]
Hetero-X test:     F(44,112) =  0.80959 [0.7843]
RESET23 test:      F(2,146)  =   2.3183 [0.1021]

COMFAC Wald test table, COMFAC F(4,148) = 21.0177 [0.0000] **
Order   Cumulative tests                 Incremental tests
  2     Chi^2(2) = 0.39429 [0.8211]      Chi^2(2) = 0.39429 [0.8211]
  1     Chi^2(4) =  84.071 [0.0000]**    Chi^2(2) =  83.676 [0.0000]**

Test for excluding:
[0] = OUTPUT
[1] = OUTPUT_1
Subset F(2,146) = 0.25212 [0.7775]

Test for linear restrictions (Rb=r):
R matrix CONS_1      CONS_2       INC      INC_1      INC_2
          1.0000      1.0000     1.0000     1.0000     1.0000
          INFLAT    INFLAT_1    INFLAT_2   Constant
         0.00000     0.00000     0.00000    0.00000
r vector
       1.0000
LinRes F(1,148) = 4.71427 [0.0315] *
```

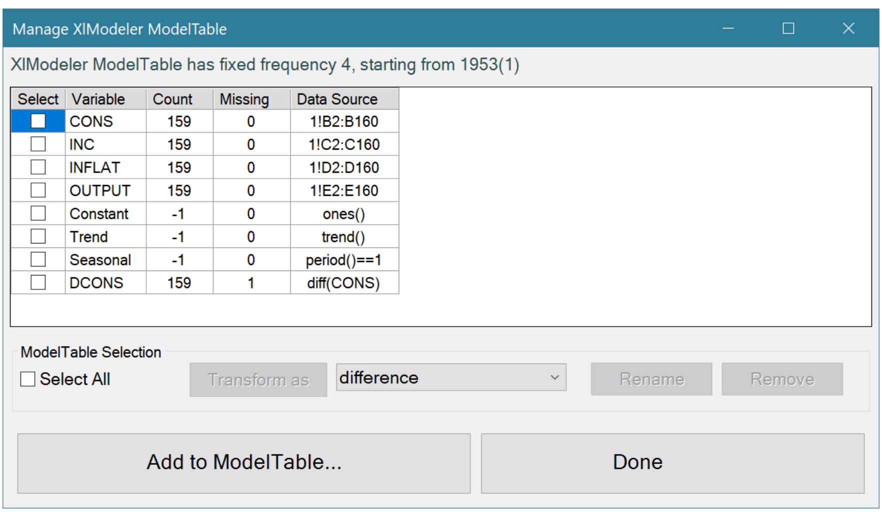
[1]This result is obtained by adding OUTPUT and OUTPUT_1 to the model, and then testing the exclusion restriction.

3.5 Sequential simplification

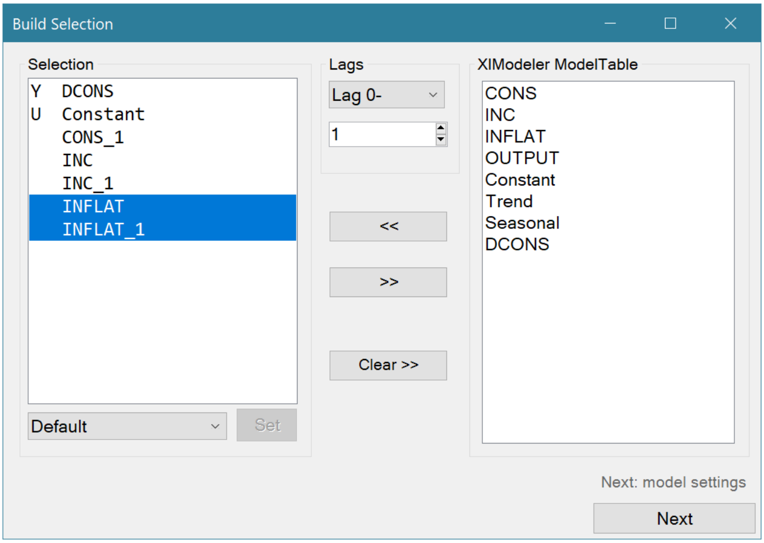
For comparability with later models, transform the dependent variable to DCONS = ΔCONS. This can be done directly in the ModelTable, by selecting DCONS, the Difference transformation:



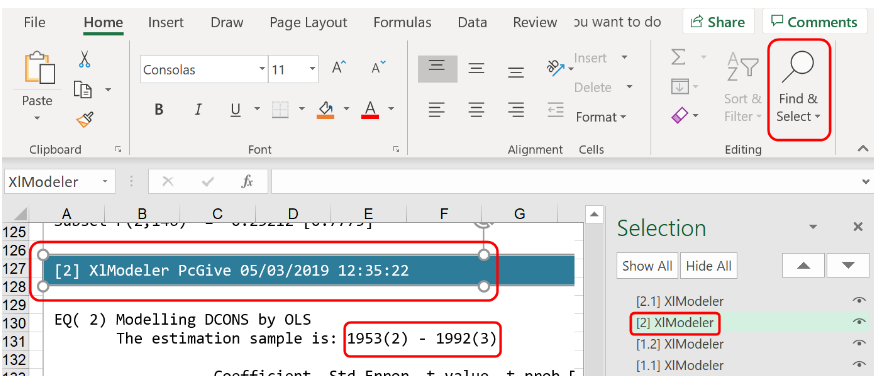
Then, after pressing Transform, the DCONS variable is created in the ModelTable:



Perhaps the quickest way to build the new model is to clear the existing model, starting from scratch. You can also remove CONS from the existing model, add DCONS and give it Y status, etc.). The model should look like this:



Repeat estimation, keeping the sample starting point to 1953(3) to match the initial model. When we first did this, we forgot to drop the first observation. If you also did that, before continuing, you may wish to delete the erroneous output. Each output panel is a 'shape', and can be deleted as follows.



Use Home/Find & Select to show the Selection Pane. The most recent output shapes are at the top. Select the second, press delete - this removes the header. Now select the top one and delete.

Using the correct sample we find:

```
EQ( 2) Modelling DCONS by OLS
       The estimation sample is: 1953(3) - 1992(3)
                  Coefficient  Std.Error  t-value  t-prob Part.R^2
CONS_1              -0.202149    0.02725    -7.42   0.000   0.2670
INC                  0.500235    0.02857     17.5   0.000   0.6700
INC_1               -0.277320    0.03808    -7.28   0.000   0.2599
INFLAT              -0.784047     0.1857    -4.22   0.000   0.1056
INFLAT_1            -0.262993     0.2057    -1.28   0.203   0.0107
Constant    U        -19.9390      8.584    -2.32   0.022   0.0345
```

```
sigma                    1.08126  RSS                   176.538482
R^2                     0.765599  F(5,151) =    98.64 [0.000]**
Adj.R^2                 0.757838  log-likelihood          -231.981
no. of observations          157  no. of parameters              6
mean(DCONS)            -0.189886  se(DCONS)                2.19724
```

Eventhough the sigma and RSS have hardly changed from the previous model, R^2 is substantially lower because of the change in the dependent variable from CONS to DCONS (i.e. subtracting lagged CONS on both sides of the equation).

3.6 Autometrics

The starting point for *Autometrics* is a model formulated in the normal way. This initial model is called the *general unrestricted model* or GUM. It should be a well-specified model, able to capture the salient features of the dependent variable and pass all diagnostic tests. Following the GUM, the main decision is the significance level for reduction. This determines at what significance regressors are removed. It also specifies the extent to which we accept a deterioration in information relative to the GUM.

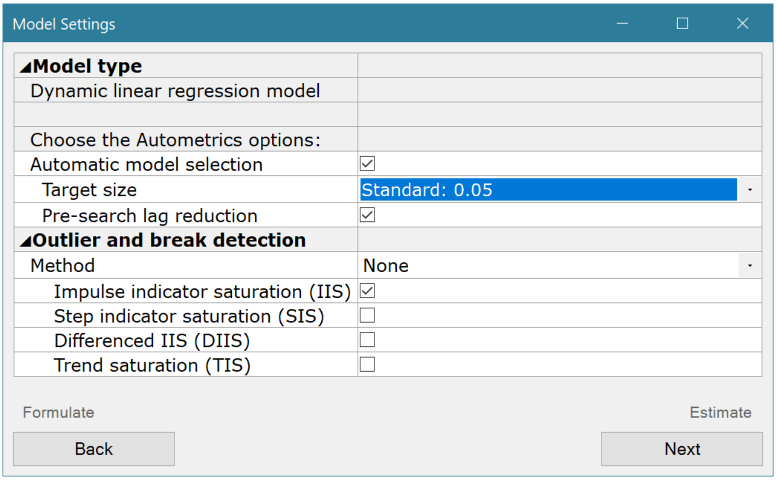
3.7 Modelling CONS

In §3.3 we shied away from using 5 lags to keep the model simple. Now we can be more ambitious, allowing for lags up to 5. Since the data is quarterly, we also add seasonals. In the previous chapter it was noted that OUTPUT should not matter. To investigate this, we add it to the model as well. Finally, we add a trend. Note that the Constant is marked as U, which means it will always be forced into the model. If you want it to be a candidate for removal, clear its status.

Accept, then mark *Autometrics*, which activates the remainder of the dialog:

- Target size
  This is the significance level that is used for reduction. Change this to 0.05.
- Pre-search lag reduction
  Pre-search lag reduction is switched on by default.
- Outlier and break detection
  Keep this at None. Alternative options are: Large residuals to automatically create dummies for large residuals in the GUM, and various forms of saturation estimation, which creates dummies for all observations.

Keep the remaining settings as shown here:



Press Next, and again in the next dialog which is unchanged from before (check that you're using the full sample). Automatic model selection is quick, but generates more output.

- **[0.1]** Initial GUM

  First the GUM is printed. The output below is not how XlModeler shows it. Instead we have sorted the regressors by t-prob, i.e. by decreasing significance:

```
GUM( 1) Modelling CONS by OLS
        The dataset is: .\OxMetrics8\data\data.in7
        The estimation sample is: 1954(2) - 1992(3)
```

		Coefficient	Std.Error	t-value	t-prob	Part.R^2
CONS_1		0.833937	0.08599	9.70	0.0000	0.4274
INC		0.511109	0.03237	15.8	0.0000	0.6643
INC_1		-0.312776	0.05848	-5.35	0.0000	0.1850
INFLAT		-0.766260	0.3703	-2.07	0.0405	0.0329
INC_3		0.120765	0.06381	1.89	0.0607	0.0276
CONS_4		0.197144	0.1103	1.79	0.0762	0.0247
CONS_5		-0.132990	0.07766	-1.71	0.0893	0.0227
CONS_3		-0.150692	0.1090	-1.38	0.1693	0.0149
OUTPUT_5		-0.0542604	0.04381	-1.24	0.2178	0.0120
INFLAT_4		0.675374	0.5832	1.16	0.2490	0.0105
OUTPUT_3		0.0514746	0.05177	0.994	0.3220	0.0078
Constant	U	-20.7235	22.88	-0.906	0.3667	0.0065
Seasonal		0.234970	0.2676	0.878	0.3815	0.0061
INFLAT_2		-0.511191	0.6022	-0.849	0.3975	0.0057
INC_2		-0.0507675	0.06353	-0.799	0.4258	0.0050
Seasonal_1		0.213602	0.2768	0.772	0.4418	0.0047
OUTPUT_4		0.0398924	0.05253	0.759	0.4490	0.0046
Seasonal_2		0.190405	0.2738	0.696	0.4880	0.0038
INC_4		-0.0396864	0.06536	-0.607	0.5448	0.0029
INFLAT_3		-0.332742	0.5959	-0.558	0.5776	0.0025
OUTPUT_2		-0.0247968	0.05115	-0.485	0.6287	0.0019
CONS_2		0.0497580	0.1101	0.452	0.6522	0.0016
INFLAT_5		-0.0770788	0.3298	-0.234	0.8156	0.0004
OUTPUT		-0.00695063	0.03279	-0.212	0.8324	0.0004
INFLAT_1		0.113980	0.5978	0.191	0.8491	0.0003

```
Trend            -0.000643632   0.004692   -0.137  0.8911   0.0001
OUTPUT_1          -0.00637418    0.05040   -0.126  0.8996   0.0001
INC_5             -0.00320849    0.05006  -0.0641  0.9490   0.0000

sigma                 1.08642  RSS                 148.719356
R^2                  0.994754  F(27,126) =   884.8 [0.000]**
Adj.R^2              0.993629  log-likelihood        -215.83
no. of observations       154  no. of parameters          28
mean(CONS)            875.591  se(CONS)              13.6115

AR 1-5 test:    F(5,121) = 0.66157 [0.6533]
ARCH 1-4 test:  F(4,146) =  1.6523 [0.1643]
Normality test: Chi^2(2) =  2.3692 [0.3059]
Hetero test:    F(51,102)= 0.97127 [0.5369]
Chow test:      F(45,81) =  1.2000 [0.2354] for break after 1981(2)
```

The first four form the core model we have worked with so far. The remaining 24 appear to be insignificant at 5%, but some may well survive into the final model. The GUM is followed by the output of the diagnostic tests that are used by *Autometrics*.

- Dimensions
  Next is some information regarding the size of the problem:

```
---------- Autometrics: dimensions of initial GUM ----------
no. of observations       154  no. of parameters          28
no. free regressors (k1)   28  no. free components (k2)    0
no. of equations            1  no. diagnostic tests        5
```

- **[0.2]** Pre-search lag reduction
  The first stage of the automatic model selection is the pre-search lag reduction:

```
[0.2] Presearch reduction of initial GUM

Starting closed lag reduction at 0.33365
Removing lags(#regressors): none

Starting common lag reduction at 0.33365
Removing lags(#regressors): 2-2(4)

Starting common lag reduction at 0.33365 (excluding lagged y's)
Removing lags(#regressors): 5-5(3) 4-4(3)

Presearch reduction in opposite order

Starting common lag reduction at 0.33365 (excluding lagged y's)
Removing lags(#regressors): 4-4(3) 2-2(3) 5-5(3)

Starting common lag reduction at 0.33365
Removing lags(#regressors): 2-2(1)

Starting closed lag reduction at 0.33365
Removing lags(#regressors): none

Encompassing test against initial GUM (iGUM) removes: none

Presearch reduction: 10 removed,  LRF_iGUM(10) [0.8430]
```

```
Presearch removed: CONS_2 INC_2 INC_4 INC_5 INFLAT_2
    INFLAT_4 INFLAT_5 OUTPUT_2 OUTPUT_4 OUTPUT_5
```

The pre-search lag reduction is done in two sequences. Only lags that are insignificant in both (at a reduced level) are removed from the initial GUM. All tests are F-tests, derived from the likelihood-ratio (LR) test — they are the standard F-tests. Lag 2, which has four variables (seasonals are not treated as lags), is the least significant with a p-value of 79%. In this case, both sequences remove exactly the same regressors, so 10 terms are removed in the pre-search, leaving 18 coefficients.

- **[0.3]** Test for empty model
  The first step after pre-search is to test for an empty model at reduced significance, which is strongly rejected:

```
[0.3] Testing GUM 0:  LRF(17) [0.0000] kept
```

- **[1.0]** Start of Autometrics tree search
  - Searching from GUM 0 The first iteration of *Autometrics* finds just one candidate models (it is more common to find multiple candidate models - try it with a free Constant to find two):

```
Searching from GUM 0  k=  17 loglik=    -219.183
Found new terminal 1  k=   6 loglik=    -222.374 SC=    3.1169

Searching for contrasting terminals in terminal paths

Encompassing test against GUM 0 removes: none

p-values in GUM 1 and saved terminal candidate model(s)
                     GUM 1  terminal 1
CONS_1          0.00000000  0.00000000
CONS_4          0.02115435  0.02115435
CONS_5          0.00183605  0.00183605
INC             0.00000000  0.00000000
INC_1           0.00000000  0.00000000
INFLAT          0.00000000  0.00000000
k                        6           6
parameters               7           7
loglik             -222.37     -222.37
AIC                 2.9789      2.9789
HQ                  3.0350      3.0350
SC                  3.1169      3.1169
```

  - Searching from GUM 1, termination
    GUM 1 is the starting point for the next search. This does not produce any new terminal candidates:

```
Searching from GUM 1  k=6 loglik=-222.374 LRF_GUM0(11) [0.8851]
Recalling terminal 1  k=6 loglik=-222.374 SC=    3.1169

Searching for contrasting terminals in terminal paths
```

- **[2.0]** Selection of final model from terminal candidates: terminal 1
  Because there were no new models when searching from GUM 1, the table headed

'p-values in Final GUM and terminal model(s)' is the same as that reported after searching from GUM 0, except that now the column of terminal one is marked. The selected model has the lowest Schwarz Criterion (SC), which is terminal 1 here. Before printing the final model, the output includes the coefficients, diagnostic tests and a summary of the search:

```
coefficients and diagnostic p-values in Final GUM and terminal model(s)
                 Final GUM  terminal 1
CONS_1             0.80973     0.80973
CONS_4             0.11861     0.11861
CONS_5            -0.12922    -0.12922
INC                0.50414     0.50414
INC_1             -0.28567    -0.28567
INFLAT             -1.0022     -1.0022
k                        6           6
parameters               7           7
loglik             -222.37     -222.37
sigma               1.0495      1.0495
AR(5)              0.80906     0.80906
ARCH(4)            0.30509     0.30509
Normality          0.60556     0.60556
Hetero             0.65103     0.65103
Chow(70%)          0.19159     0.19159
                               =======

p-values of diagnostic checks for model validity
          Initial GUM     cut-off   Final GUM     cut-off Final model
AR(5)         0.65329     0.01000     0.80906     0.01000     0.80906
ARCH(4)       0.16430     0.01000     0.30509     0.01000     0.30509
Normality     0.30587     0.01000     0.60556     0.01000     0.60556
Hetero        0.53688     0.01000     0.65103     0.01000     0.65103
Chow(70%)     0.23543     0.01000     0.19159     0.01000     0.19159


Summary of Autometrics search
initial search space     2^27  final search space        2^6
no. estimated models       29  no. terminal models         1
test form                LR-F  target size      Standard:0.05
large residuals            no  presearch reduction      lags
backtesting              GUM0  tie-breaker                SC
diagnostics p-value      0.01  search effort        standard
time                     0.02  Autometrics version       2.0
```

The final model, which differs from the one of the previous chapter, has lags four and five of CONS (with almost opposite coefficients), as additional variables. Given 28 variables in the GUM, at 5% significance one might expect one or two to be retained by chance; a 1% level would reduce that 'spurious' retention rate to about one variable every three times that such a selection exercise was conducted. Any actions on the Test menu now relate to this model. For example, testing that the coefficients on CONS_4 and CONS_5 sum to zero using Test/Linear Restrictions is accepted with a p-value of 63%. Therefore, the long-run is not greatly changed from that reported in §3.3:

```
Solved static long-run equation for CONS
                    Coefficient  Std.Error  t-value  t-prob
```

```
INC                      1.08753    0.04200     25.9   0.0000
INFLAT                  -4.98868     0.5069    -9.84   0.0000
Constant                -85.3400      37.66    -2.27   0.0249
Long-run sigma = 5.22436
```

3.8 DHSY revisited

The DHSY model (Davidson, Hendry, Srba, and Yeo, 1978) is an equilibrium-correction model for the logarithm of consumption, c_t, where the equilibrium correction is the gap between consumption and income, y_t, with an additional price term, $\Delta_4 p_t$. The DHSY model is seasonal: it uses fourth differences (the data are quarterly), and the equilibrium is towards the gap from a year ago. There is a dummy, D_{budget}, for budget effects in 1968: $+1$ in 1968(1) and -1 in 1968(2), and a dummy, D_{VAT}, for the introduction of VAT: $+1$ in 1973(1) and -1 in 1973(2). DHSY use $DV_t = D_{\text{budget}} + D_{\text{VAT}}$ in their model.

The data set is provided as `DHSY.xlsx` with the ModelTable already constructed:

	A	B	C	D	E	F
1	**ModelTable**					
2	Frequency	4				
3	Start Year	1957				
4	Start Period	1				
5	Dates Style	Fixed Frequency				
6	Dates Source	DHSY!A2:A83				
7						
8	**Variables**					
9	**Name**	**Data Source**	**Count**	**Missing**		
10	LY	DHSY!B2:B83	82	4		
11	LC	DHSY!C2:C83	82	4		
12	D4LPC	DHSY!D2:D83	82	8		
13	D6812	DHSY!E2:E83	82	0		
14	D7312	DHSY!F2:F83	82	0		
15	DV	DHSY!G2:G83	82	0		
16	ECM	DHSY!H2:H83	82	8		
17	LC-LY	DHSY!I2:I83	82	4		
18	D4LY	diff(LY,4)	82	8		
19	D4LC	diff(LC,4)	82	8		
20	DD4LPC	diff(D4LPC)	82	9		
21	D4DV	diff(DV,4)	82	4		
22	Constant	ones()	-1	0		
23	Trend	trend()	-1	0		
24	Seasonal	period()==1	-1	0		
25	DD4LY	diff(D4LY)	-1	0		
26						

Here LC is c_t, LY is y_t and D4LPC is $\Delta_4 p_t$. This allows us to formulate their model. Estimation is over the sample period 1959(2)–1975(4). Using Further Output/LaTeX format:

$$\begin{aligned}\text{D4LC} \;=\; & -\underset{(0.012)}{0.093}\;\text{LC-LY}_{t-4} + \underset{(0.029)}{0.48}\;\text{D4LY}_t - \underset{(0.023)}{0.12}\;\text{D4LPC}_t \\ & -\underset{(0.04)}{0.23}\;\text{DD4LY}_t - \underset{(0.1)}{0.31}\;\text{DD4LPC}_t + \underset{(0.0022)}{0.0065}\;\text{D4DV}_t\end{aligned}$$

With standard errors in parentheses.

To use Autometrics, formulate the GUM with LC as the dependent variable, and up to 5 lags of LC, LY and D4LPC as explanatory variables. The constant is already there, but its status needs to be changed from U to a normal regressor. Add the differenced dummy D4DV without further lags, and finally add the seasonals. Estimation is over the sample period 1959(2)–1975(4).

Autometrics at 2.5% with pre-search lag reduction finds three terminal models:

```
p-values in Final GUM and terminal model(s)
                 Final GUM  terminal 1  terminal 2  terminal 3
LC_4            0.00000000  0.00000000  0.00000000  0.00000000
LY              0.00000000  0.00000000  0.00000000  0.00000000
LY_1            0.00000984  0.00001130  0.00003436  0.00000567
LY_4            0.04766878  0.00481955           .           .
LY_5            0.00000844  0.00000006  0.00000027  0.00000000
D4LPC           0.00013293  0.00020981  0.00000000  0.00001474
D4LPC_1         0.03291580  0.01494571           .  0.00933580
D4DV            0.00163508  0.00201634  0.00034527  0.00027128
Seasonal        0.08409064           .  0.04147579           .
Seasonal_1      0.02567385           .  0.00071121  0.01432980
Seasonal_2      0.15804560           .  0.07652718           .
k                       11           8           9           8
parameters              11           8           9           8
loglik              256.60      253.24      250.83      252.12
AIC                -7.3312     -7.3205     -7.2189     -7.2872
HQ                 -7.1880     -7.2163     -7.1018     -7.1830
SC                 -6.9693     -7.0573     -6.9228     -7.0239
                               =======
```

These are all statistically valid reductions of the GUM. When there is a tie like this, the final model is chosen by the Schwarz Criterion (SC). The selected model is very similar to that found by Davidson, Hendry, Srba, and Yeo (1978) (after much less effort than theirs!):

$$\begin{aligned}
\mathrm{LC} = &\underset{(0.028)}{0.92}\ \mathrm{LC}_{t-4} + \underset{(0.038)}{0.27}\ \mathrm{LY}_t + \underset{(0.04)}{0.19}\ \mathrm{LY}_{t-1} \\
&- \underset{(0.046)}{0.14}\ \mathrm{LY}_{t-4} - \underset{(0.04)}{0.25}\ \mathrm{LY}_{t-5} - \underset{(0.095)}{0.38}\ \mathrm{D4LPC}_t \\
&+ \underset{(0.1)}{0.25}\ \mathrm{D4LPC}_{t-1} + \underset{(0.0021)}{0.0069}\ \mathrm{D4DV}_t
\end{aligned}$$

The results of this chapter have been saved in `data-2.xlsx` and `DHSY.xlsx`, both stored in your user folder under XlModeler. This finishes our coverage of regression models in XlModeler. XlModeler is based on PcGive, and Hendry and Doornik (2013) provides a full documentation of the output of the program.

In the next chapter we turn to financial econometrics, considering volatility models.

Chapter 4

Volatility Model

4.1 Getting started

The data set for this chapter is the daily Nasdaq stock index (IXIC). The file `Nasdaq.xlsx` is provided in your user folder under `XlModeler\data`. This data set was downloaded from Yahoo! Finance, and, if you wish, you can follow along with a newer sample.

The downloaded data consists of the Date, Open, High, Low, Close, Adj Close, and Volume. This is what it looks like on our computer:

	A	B	C	D	E	F	G	H	I
1	Date	Open	High	Low	Close	Adj Close	Volume	Monday	Friday
2	03/01/2005	2184.75	2191.6	2148.72	2152.15	2152.15	2.19E+09	1	0
3	04/01/2005	2158.31	2159.64	2100.56	2107.86	2107.86	2.69E+09	0	0
4	05/01/2005	2102.9	2116.75	2091.24	2091.24	2091.24	2.38E+09	0	0
5	06/01/2005	2098.51	2103.9	2088.03	2090	2090	2.17E+09	0	0
6	07/01/2005	2099.95	2103.39	2076.69	2088.61	2088.61	2.19E+09	0	1
7	10/01/2005	2087.62	2111.43	2086.66	2097.04	2097.04	2.1E+09	1	0
8	11/01/2005	2089.07	2090.62	2072.62	2079.62	2079.62	2.21E+09	0	0
9	12/01/2005	2089.7	2093.44	2066.79	2092.53	2092.53	2.26E+09	0	0
10	13/01/2005	2093.54	2094.8	2067.94	2070.56	2070.56	2.11E+09	0	0

The sample runs from 3 January 2005 to 31 December 2018. There are no observations for weekends and holidays in which the market is closed. We shall use the Adjusted close, which is price adjusted for both dividends and splits.

Two additional variables have been created in the spreadsheet: a Monday dummy and a Friday dummy. If you download your own data set, this can be done as follows:

1. in cell H2 enter

```
=IF(WEEKDAY(A2)=2,1,0)
```

2. Press Ctrl+Shift+End to extend the selection to the last cell and fill the column with the formula (just scrolling down with so many observations will take forever). Finally, give the column a name.

The formula for Friday is =IF(WEEKDAY(A2)=6,1,0).

The first step in using XlModeler is to create the ModelTable, as discussed in Chapter 1. However, this has already been done in Nasdaq.xlsx:

	A	B	C	D	E
1	**ModelTable**				
2	Frequency	1			
3	Start Year	1			
4	Start Period	1			
5	Dates Style	Calendar			
6	Dates Source	^IXIC!A2:A3524			
7					
8	**Variables**				
9	**Name**	**Data Source**	**Count**	**Missing**	
10	Adj Close	^IXIC!F2:F3524	3523	0	
11	Monday	^IXIC!H2:H3524	3523	0	
12	Friday	^IXIC!I2:I3524	3523	0	
13	Constant	ones()	-1	0	
14	Trend	trend()	-1	0	
15	ixic	100*dlog("Adj Close")	3523	1	
16					

^IXIC | XlModeler.Table

The Constant and the Trend are automatically created by XlModeler, but not used in volatility models.

The object of interest is the daily return (in %), which are defined as

$$y_t = 100\left[log(p_t) - log(p_{t-1})\right], \tag{4.1}$$

where p_t is the price series at time t. The returns are called the growth rate in the ModelTable dialog, with transformation 100*dlog("Adj Close"). We renamed this to ixic (in lowercase).

Whereas the ModelTable shown in §1.1 and used in previous chapters has a fixed frequency, the current table has dates of type Calendar. This associates a date with every observation.

4.2 Preliminary analysis

Before delving into a volatility model, we can use the regression model to do our preliminary analysis. Regressing ixic on just a Constant shows:

```
EQ( 1) Modelling ixic by OLS
       The estimation sample is: 2005-01-04 - 2018-12-31

                  Coefficient  Std.Error  t-value  t-prob Part.R^2
Constant       U    0.0319686    0.02167     1.48  0.1402   0.0006
```

```
sigma                    1.28603   RSS                    5823.301
log-likelihood             -5883
no. of observations         3522   no. of parameters             1
mean(ixic)             0.0319686   se(ixic)                1.28603

AR 1-2 test:       F(2,3519) =    11.691 [0.0000]**
ARCH 1-1 test:     F(1,3520) =    204.12 [0.0000]**
Normality test:    Chi^2(2)  =    2382.8 [0.0000]**
```

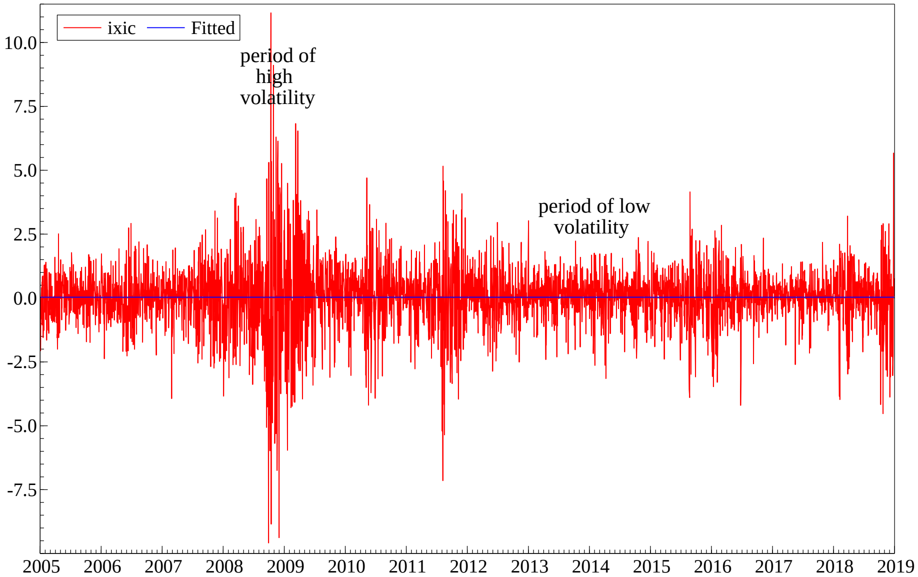


Figure 4.1 NASDAQ returns

Graphic analysis can show us the ‘actual and fitted’ with the latter just being the intercept, Figure 4.1. The NASDAQ returns variable exhibits volatility clustering as periods of low volatility mingle with periods of high volatility. This is a clear sign of presence of ARCH effect in the series. We also see from the ARCH mis-specification test that the absence of ARCH is rejected.

4.2.1 Nonnormality

Figure 4.2 shows that the unconditional distribution of the NASDAQ returns (the thicker line) is not normally distributed: it is more peaked than the normal density with the same mean and variance (this is the dashed line) and has fatter tails. The inset panels of the left and right tail emphasize this. They also indicate that the left tail is somewhat fatter than the right, so large negative returns seem a bit more likely than large positive ones. The normality test also indicates the departure from normality of the NASDAQ returns. We can look at this in more detail using More/Test and selecting Normality test:

```
Normality test for residuals
Observations                3522
Mean                     0.00000
Std.Devn.                 1.2858
Skewness                -0.28344
Excess Kurtosis           7.5342
Minimum                  -9.6197
Maximum                   11.127
Median                  0.055229
Madn                     0.85131
Asymptotic test:   Chi^2(2)   =    8377.4 [0.0000]**
Normality test:    Chi^2(2)   =    2382.8 [0.0000]**
```

Recall that the skewness coefficient (SK) equals 0 for a symmetric distribution while the kurtosis coefficient (KU) equals 3 for the normal distribution. The excess kurtosis equals $KU - 3$. Formally these moments are expressed as

$$SK = \frac{E[(y-\mu)^3]}{\sigma^3} \text{ and } KU = \frac{E[(y-\mu)^4]}{\sigma^4}$$

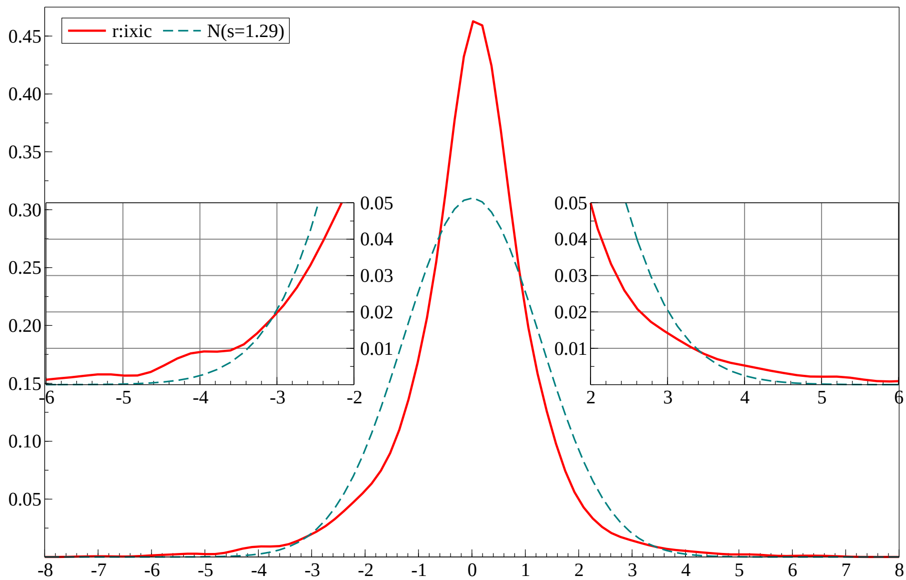


Figure 4.2 Distribution of NASDAQ returns

4.2.2 Autocorrelations

When the error term is not independent of previous errors, it is said to be autocorrelated. Autocorrelation of order h is computed as follows

$$r_h = \frac{cov(y_t, y_{t-h})}{\sigma_{y,t}\sigma_{y,t-h}}.$$

Plotting the sample autocorrelations against the lag gives a first visual impression of the magnitude of the autocorrelation problem. This plot is called “autocorrelogram” or ACF and it provides information about simple correlation coefficients. When the errors are strongly time dependent, the autocorrelations tend to be fairly large even for large values of h. We can use the Graphic Analysis to plot tha ACF of the residuals (which are just the returns in this case), as well as of the squared returns.

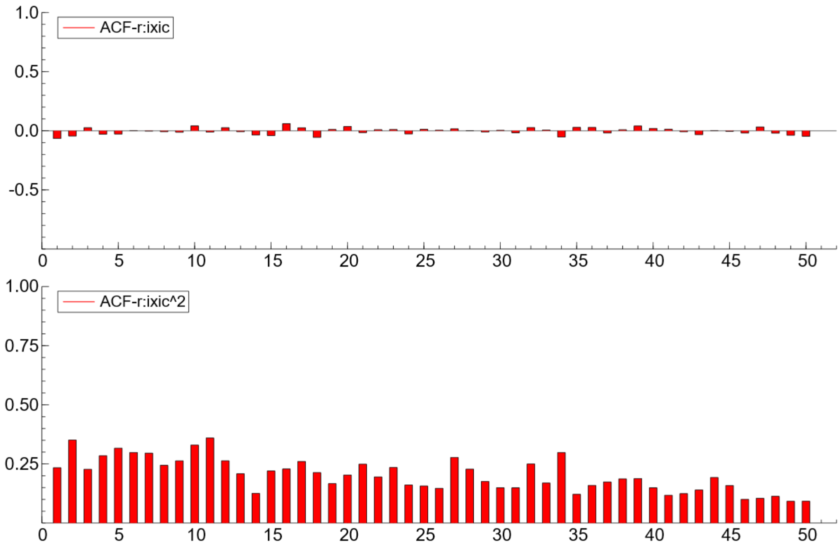


Figure 4.3 Autocorrelations of NASDAQ returns

The top panel of Figure 4.3 suggests that the daily return series of the NASDAQ is a short memory process (in the level) but that an AR(1) term might be needed in the conditional mean equation.

We have previously seen from the visual inspection that the NASDAQ exhibits volatility clustering as periods of low volatility alternate with periods of high volatility. This is reflected in the strong autocorrelation of the squared returns, see the bottom panel of Figure 4.3. The aim of a volatility model is to capture this deature of the data — a normal regression model cannot do this.

4.3 The ARCH model

Our starting point is a univariate time series y_t. If Ω_{t-1} is the information set at time $t-1$, we can define its functional form as:

$$y_t = E(y_t|\Omega_{t-1}) + \varepsilon_t, \tag{4.2}$$

where $E(.|.)$ denotes the conditional expectation operator and ε_t is the disturbance term (or unpredictable part), with $E(\varepsilon_t) = 0$ and $E(\varepsilon_t \varepsilon_s) = 0, \forall\, t \neq s$.

In the regression models of the previous chapters we specified the expectation as $x_t'\delta$, and assumed, for inference purposes, that the error term had an $N[0, \sigma^2]$ distribution. However, Figure 4.2 and also the ACF of squared residuals in Figure 4.3 shows that this assumption of constant variance does not hold for the Nasdaq returns. This is the case in general for financial time-series, which led to the development of ARCH-style and stochastic volatility models.

4.3.1 The volatility equation

More than three decades ago, Engle (1982) introduced the Autoregressive Conditional Heteroscedasticity (ARCH) model:

$$\varepsilon_t = z_t \sigma_t \quad z_t \sim i.i.d.N[0,1] \tag{4.3}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2. \tag{4.4}$$

All volatility models offered by XlModeler are ARCH-type models of the form $\varepsilon_t = z_t \sigma_t$, but with many different options for the precise form of (4.4) and the distribution of z_t.

The ARCH model can describe volatility clustering. The conditional variance of ε_t is an increasing function of the square of the shock that occurred at $t-1$. Consequently, if this was large in absolute value, σ_t^2 and thus ε_t is expected to be large (in absolute value) as well. Note that, even if the conditional variance of an ARCH model is time-varying, i.e. $\sigma_t^2 = E(\varepsilon_t^2|\Omega_{t-1})$, the unconditional variance of ε_t is constant and, provided that $\omega > 0$ and $\alpha_1 + ... + \alpha_q < 1$, we have:

$$\sigma^2 \equiv E[E(\varepsilon_t^2|\Omega_{t-1})] = \frac{\omega}{1 - \sum_{i=1}^{q} \alpha_i}. \tag{4.5}$$

When z_t is normally distributed, $E(z_t^3) = 0$ and $E(z_t^4) = 3$. Consequently, $E(\varepsilon_t^3) = 0$ and the skewness of y is zero. The kurtosis coefficient for the ARCH(1) process is $3(1-\alpha_1^2)/(1-3\alpha_1^2)$ if $\alpha_1 < \sqrt{1/3} \approx 0.577$. In this case, the unconditional distribution has fat tails whenever $\alpha_1 > 0$. In most applications, the excess kurtosis implied by the normal ARCH model is not enough to mimic the features of observed financial data.

Positivity Constraints The σ_t^2 terms must be positive for all t. Sufficient conditions to ensure that the conditional variance in Equation (4.4) is positive are given by $\omega > 0$ and $\alpha_i \geq 0$.

Variance Targeting Variance targeting in the ARCH(1) model replaces ω by $\sigma^2(1 - \alpha_1)$. The unconditional variance σ^2 can be estimated consistently by its sample counterpart.

Explanatory Variables Explanatory variables can be added to the conditional variance (note the triple use of ω):

$$\omega_t = \omega + \sum_{i=1}^{n_2} \omega_i x_{i,t}. \tag{4.6}$$

This invalidates the standard conditions for positive variance.

Variance targeting is also affected, and ω is now replaced by $\sigma^2 \left(1 - \sum_{i=1}^{q} \alpha_i\right) - \sum_{i=1}^{n_2} \omega_i \bar{x}_i$, where $\bar{x}_i$ is the sample average of variable $x_{i,t}$ (assuming the stationarity of the n_2 explanatory variables). In other words, the explanatory variables are centered.

4.3.2 The mean Equation

The mean of the volatility model may have regressors, as well as autoregressive (AR) and/or moving average (MA) components:

$$\Psi\left(L\right)\left(y_t - x_t'\delta\right) = \Theta\left(L\right)\varepsilon_t, \tag{4.7}$$

where L is the lag operator such that $L^k y_t = y_{t-k}$, $\Psi\left(L\right) = 1 - \sum_{i=1}^{n} \psi_i L^i$ and $\Theta\left(L\right) = 1 + \sum_{j=1}^{s} \theta_j L^j$.

Furthermore, long memory can be captured through a fractionally integrated ARMA process:

$$\Psi\left(L\right)\left(1 - L\right)^{\zeta}\left(y_t - \mu_t\right) = \Theta\left(L\right)\varepsilon_t, \tag{4.8}$$

where the operator $(1 - L)^{\zeta}$ accounts for the long memory of the process.

Another feature is the availability of ARCH ‘in-mean’ models, where the conditional variance is an explanatory variable in the mean:

$$\mu_t = \mu + \vartheta \sigma_t^k, \tag{4.9}$$

with $k = 1$ to include the conditional standard deviation and $k = 2$ for the conditional variance. Such an ARCH-M model is often used in financial applications where the expected return on an asset is related to the expected asset risk. The estimated coefficient of the expected risk is a measure of the risk-return tradeoff.

4.3.3 Estimation

The evaluation of σ_t^2 in (4.4) depends on past (squared) residuals. These are not observed before $t = 1$, and, to initialize the process, we use their sample mean.

Estimation of ARCH-type models is commonly done by maximum likelihood requiring an additional assumption about the innovation process z_t. The default is to use a normal distribution, and Weiss (1986) and Bollerslev and Wooldridge (1992) show that the quasi-maximum likelihood (QML) estimator is then consistent if the conditional mean and the conditional variance are correctly specified. This estimator is, however, inefficient with the degree of inefficiency increasing with the degree of departure from normality (Engle and González-Rivera, 1991). It is, of course, efficient when normality holds.

Other available distributions are the Student-t distribution, the Generalized Error distribution (GED) and the skewed-Student distribution (Lambert and Laurent, 2000).

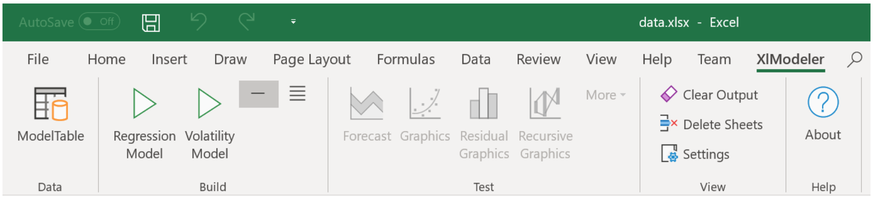
Likelihood maximization methods Three numerical methods are available to maximize the loglikelihood:

1. The quasi-Newton method of Broyden, Fletcher, Goldfarb and Shanno (BFGS).
2. A constraint optimization technique that uses the MaxSQPF algorithm. `MaxSQPF` implements a sequential quadratic programming technique to maximize a non-linear function subject to non-linear constraints, similar to Algorithm 18.7 in Nocedal and Wright (1999). `MaxSQPF` is particularly useful to impose the stationarity and/or positivity constraints like $\alpha_1 \geq 0$ in the ARCH(1) model.
3. A simulated annealing algorithm for optimizing non-smooth functions with possible multiple local maxima.

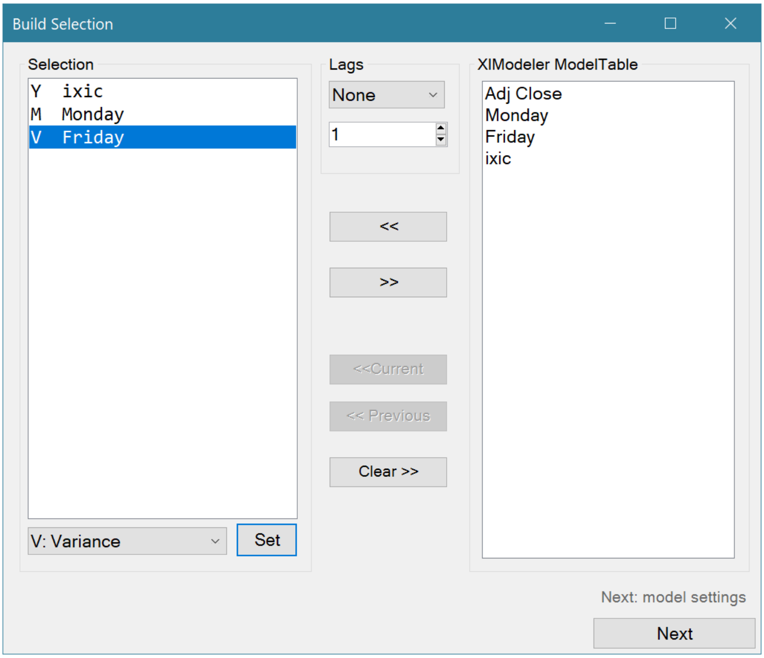
4.4 Building an ARCH volatility model

First, we consider the autoregressive conditional heteroscedasticity (ARCH) model introduced by Engle (1982). But we shall see that a generalization of this to GARCH is more effective at capturing the volatility clustering of the NASDAQ returns series. This section takes you through the steps needed to build and estimate an ARCH(1) volatility model. The next section considers many aspects of this model in detail.

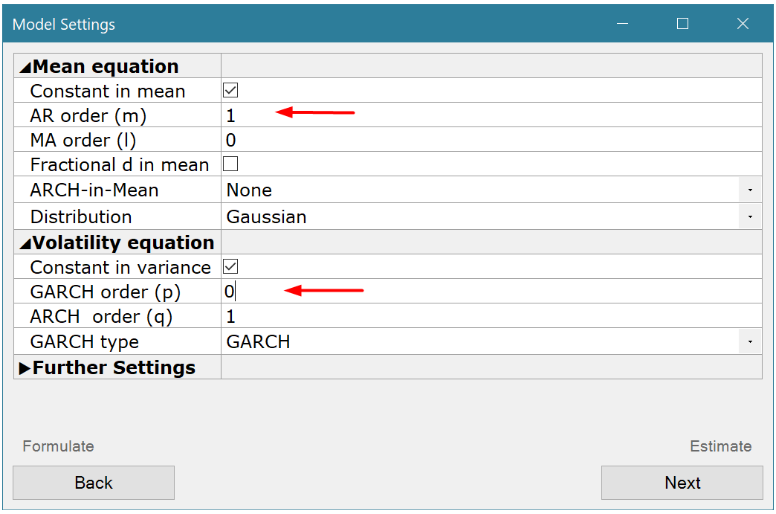
Estimating our first model for ixic is simple with XlModeler. Because the ModelTable has already been constructed, we can start by clicking on Volatility Model in the Excel ribbon bar:



A list with all the variables of the ModelTable appears in the right-hand frame. To select variables that will enter your model, click on the variable name and then click on the << button. There are three possible statuses for each variable (see the list of statuses under the Selection frame): dependent variable (Y variable), regressor in the conditional mean (Mean), or regressor in the conditional variance (Variance). In the univariate module, only one Y variable per model is accepted. However one can include several regressors in the conditional mean and the conditional variance equations. In the example the Monday dummy is included in the conditional mean and the Friday dummy in the conditional variance equation:

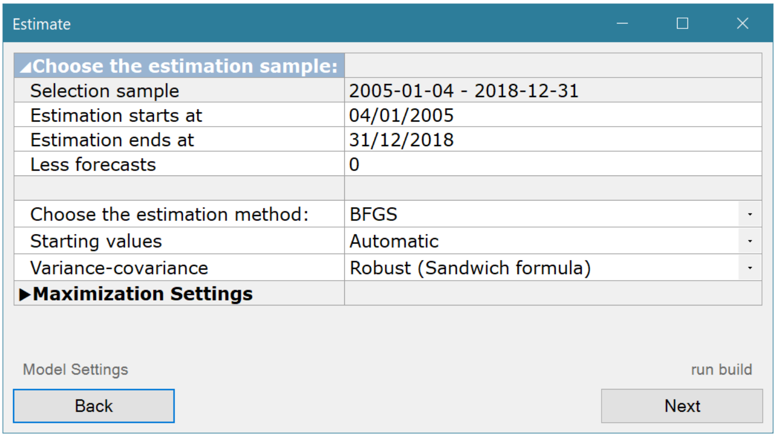


Once the Next button is pressed, the Model Settings box automatically appears. This box is to select the specification of the model: AR(FI)MA orders and the distribution for the mean equation; GARCH orders and type of GARCH model for the variance equation. The default specification is for a GARCH(1,1) model with normal errors. We are going to build an AR(1)-ARCH(1) model, in extended notation: ARMA(0,0)-GARCH(1,1). This is specified as follows:

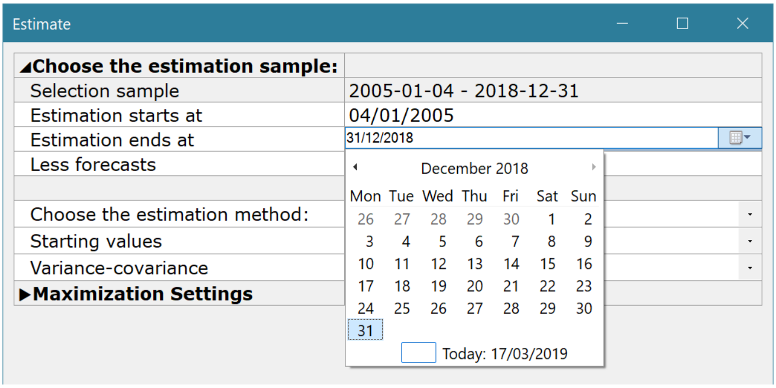


Starting Values are automatically chosen, but there is an option to enter them manually, element by element. The automatic method is obviously the easiest to use and is recommended unless there are problems with convergence.

Next, the Estimate window proposes options on two important characteristics of the model: the sample size and the estimation method:



Changing the end date is through a calendar control:



When the variable corresponding to the date is correctly formatted, the sample can conveniently be fixed based on starting and ending date. The number of forecasts can be also subtracted when out-of-sample forecasting is to be performed. For a fixed frequency ModelTable, the method of sample selection is identical to regression models.

By default, robust standard errors are reported.

The Maximization Settings relate to the output and convergence tolerances of numerical maximization procedures. All options are maintained from one build to the next.

Click Next and the estimation procedure is launched if the automatic starting values are used. Otherwise, an additional dialog box appears to modify the default starting values.

The AR(1)-ARCH(1) model that we estimated is:

$$\begin{aligned} y_t &= \rho y_{t-1} + \mu + \gamma_1 \text{Monday}_t + \varepsilon_t, \\ \varepsilon_t &= z_t \sigma_t \quad z_t \sim i.i.d.\ N[0,1], \\ \sigma_t^2 &= \omega + \gamma_2 \text{Friday}_t + \alpha_1 \varepsilon_{t-1}^2. \end{aligned}$$

The output of the estimation is given in the **XlModeler.Out** sheet in Excel:

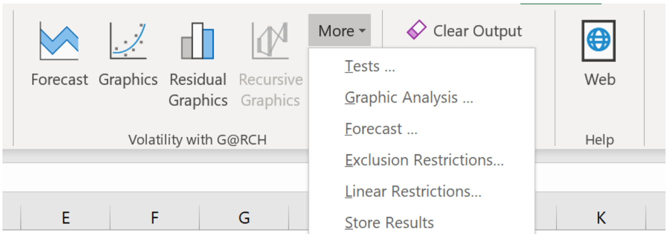
```
The estimation sample is:  2005-01-04 - 2018-12-31
The dependent variable is: ixic
Mean Equation:  ARMA (1, 0) model.
1 regressor(s) in the conditional mean.
Variance Equation:  GARCH (0, 1) model.
1 regressor(s) in the conditional variance.
Normal distribution.

Strong convergence using numerical derivatives
Log-likelihood = -5692.29
Please wait : Computing the Std Errors ...

 Robust Standard Errors (Sandwich formula)
                  Coefficient  Std.Error  t-value  t-prob
Cst(M)               0.068847   0.023222    2.965  0.0031
Monday (M)          -0.028633   0.055363  -0.5172  0.6051
AR(1)               -0.059025   0.072989  -0.8087  0.4188
Cst(V)               1.260808    0.11311    11.15  0.0000
Friday (V)          -0.364409    0.12689   -2.872  0.0041
ARCH(Alpha1)         0.287649   0.057891    4.969  0.0000

No. Observations :      3522  No. Parameters  :         6
Mean (Y)         :   0.03197  Variance (Y)    :   1.65341
Skewness (Y)     :  -0.28344  Kurtosis (Y)    :  10.53423
Log Likelihood   : -5692.286  Alpha[1]+Beta[1]:   0.28765

The sample mean of squared residuals was used to start recursion.
Positivity & stationarity constraints are not computed because there are
explanatory variables in the conditional variance equation.

Estimated Parameters Vector :
 0.068847;-0.028633;-0.059025; 1.260808;-0.364409; 0.287654
Elapsed Time : 0.181 seconds (or 0.00301667 minutes).
```

Parameters labelled '(M)' relate to the conditional mean while those labelled '(V)' relate to the conditional variance equation. AR(1) and ARCH(Alpha1) correspond to ρ and α_1, respectively.

Surprisingly, the AR(1) coefficient ρ is not significantly different from 0 (we will come back to this issue latter) while it was expected to be significantly negative. Interestingly, the returns and volatility are, on average, found to be lower on Monday and on Friday, respectively. Furthermore, the ARCH coefficient α_1 is highly significant (rejecting the null of no ARCH effects) but does not capture the kurtosis (it is below 0.577, with implied kurtosis of about 1.2). The log-likelihood value is -5703.476.

4.5 Testing the volatility model

As a next step, it is desirable to test the adequacy of this ARCH model. Several options to do so are thus available after the estimation of the model: the Forecast, Graphics and Residual Graphics buttons, as well as under the More dropdown: Test, Graphic Analysis, Forecast, Exclusion Restrictions, Linear Restrictions and Store Results:



The Graphics button shows the conditional mean and standard deviation, while the residual graphics shows the squared residuals and the distribution of the standardized residuals. The Graphic Analysis... option allows to plot different graphics:

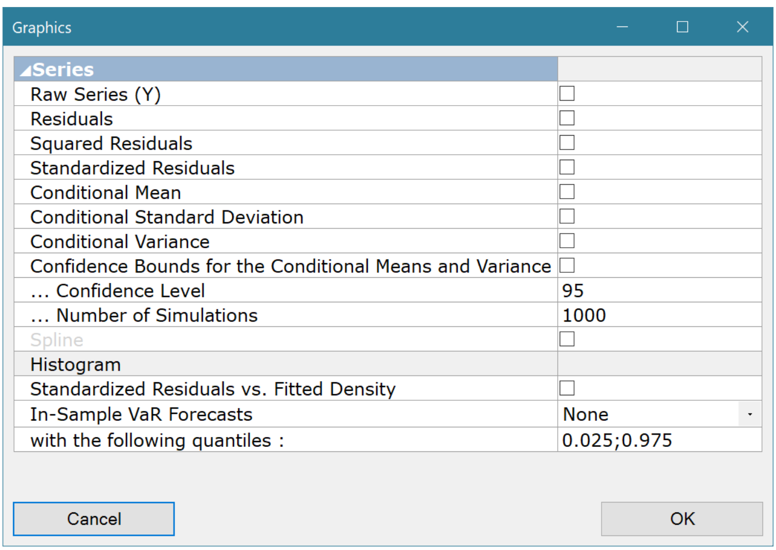


Figure 4.4 plots the conditional variance ($\hat{\sigma}_t^2$) as well as the histogram of the standardized residuals ($\hat{z}_t = \frac{\hat{\varepsilon}_t}{\hat{\sigma}_t}$) obtained with the AR(1)-ARCH(1) model, together with a kernel estimation of its unconditional distribution (solid line) and the $N(0,1)$ (dotted line).

In most applications of GARCH models, the estimated volatility is plotted without confidence bounds. However, XlModeler can provide in-sample confidence bands for the conditional mean and conditional variance of univariate GARCH-type models, based on the results of Blasques, Lasak, Koopman, and Lucas (2016). The bands are obtained by simulation. As an illustration, 95% confidence bounds for the conditional

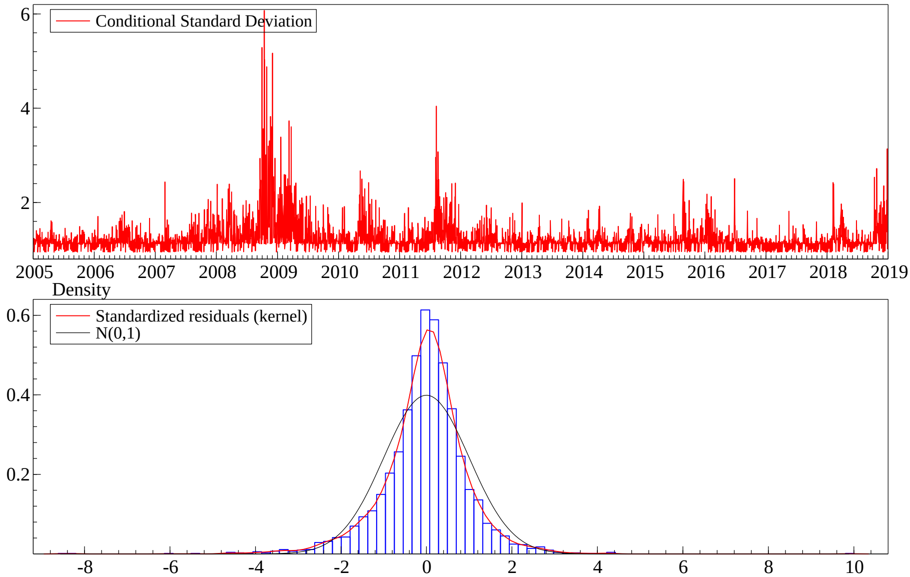


Figure 4.4 Conditional variance of the NASDAQ and estimated unconditional density of the standardized residuals

mean and the conditional variance of our model is plotted in Figure 4.5, together with the estimated conditional moments.

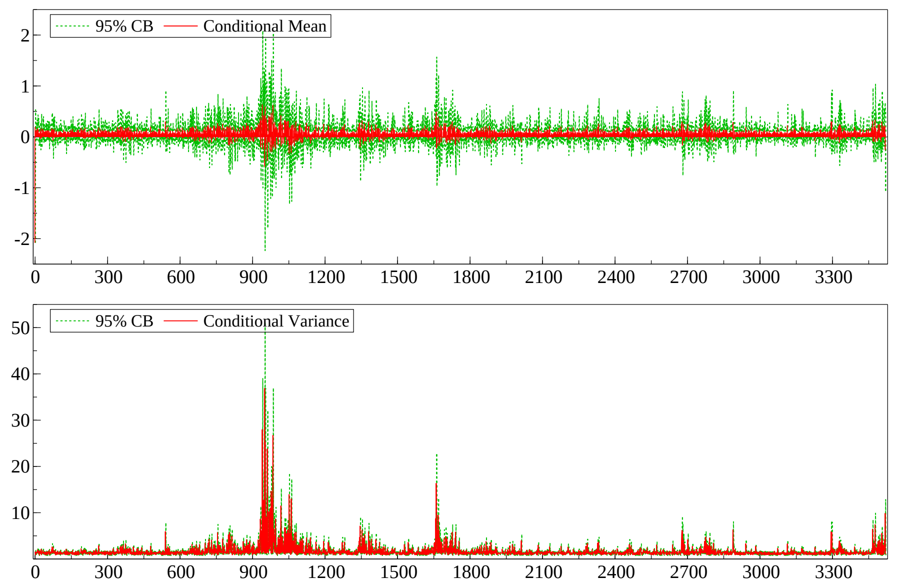
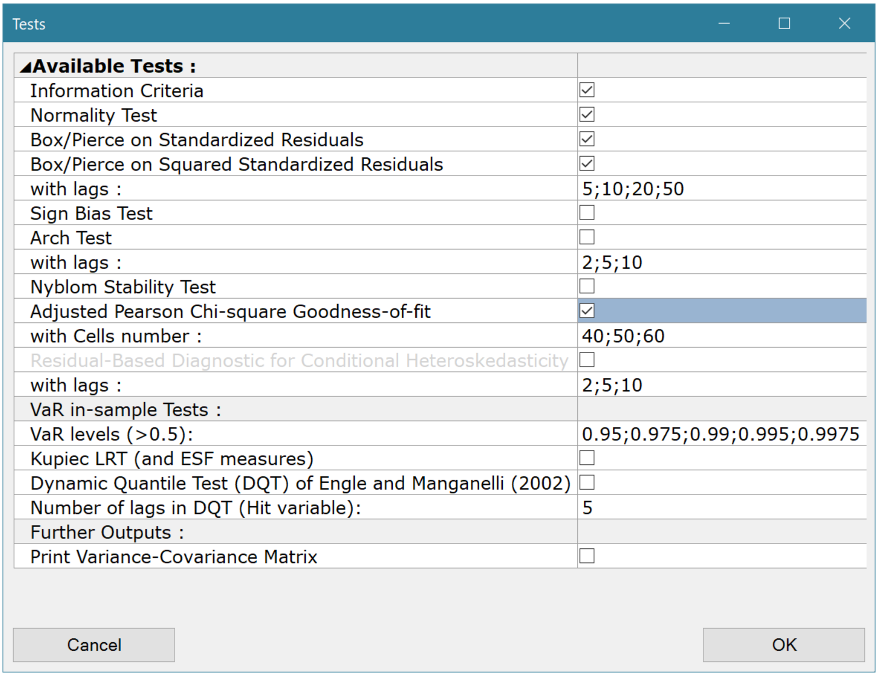


Figure 4.5 In-sample confidence bounds

4.5.1 Misspecification Tests

The Tests... option allows the user to run different tests but also to print the variance-covariance matrix of the estimated parameters. We choose the first four options, as well as the adjusted Pearson test, see the screen capture below the output.

```
TESTS :
---------
Information Criteria (to be minimized)
Akaike          3.235824  Shibata         3.235818
Schwarz         3.246330  Hannan-Quinn    3.239572
---------------

Normality Test

                   Statistic       t-Test      P-Value
Skewness            -0.49758       12.061   1.7070e-33
Excess Kurtosis       7.7770       94.278      0.00000
Jarque-Bera           9021.1         .NaN      0.00000
---------------

Q-Statistics on Standardized Residuals
  --> P-values adjusted by 1 degree(s) of freedom
  Q(  5) =  8.20137   [0.0844742]
  Q( 10) =  14.8156   [0.0961278]
  Q( 20) =  38.8667   [0.0045958]**
  Q( 50) =  96.0263   [0.0000687]**
H0 : No serial correlation
    ==> Accept H0 when prob. is High [Q < Chisq(lag)]
---------------

Q-Statistics on Squared Standardized Residuals
  --> P-values adjusted by 1 degree(s) of freedom
  Q(  5) =  320.105   [0.0000000]**
  Q( 10) =  894.033   [0.0000000]**
  Q( 20) =  1245.69   [0.0000000]**
  Q( 50) =  2337.25   [0.0000000]**
H0 : No serial correlation
    ==> Accept H0 when prob. is High [Q < Chisq(lag)]
---------------

Adjusted Pearson Chi-square Goodness-of-fit test

# Cells(g)  Statistic      P-Value(g-1)     P-Value(g-k-1)
    40      421.8955         0.000000           0.000000
    50      436.7734         0.000000           0.000000
    60      447.6422         0.000000           0.000000
Rem.: k = 6 = # estimated parameters
```



Without going too deeply into the analysis of these results, they suggest that the model does not capture the dynamics of the NASDAQ returns.

The Q-statistics on standardized and squared standardized residuals, as well as the adjusted Pearson Chi-square goodness-of-fit test (with different cell numbers) reject the null hypothesis of a correct specification. This result is not very surprising. Early empirical evidence has indeed shown that a high ARCH order has to be selected to catch the dynamics of the conditional variance (thus involving the estimation of numerous parameters). This will lead us to the GARCH model in the next chapter.

Note that the residual-based diagnostic test for conditional heteroskedasticity is disabled when robust standard errors are used. So we can change to second derivatives. This has quite an impact in the current model:

		Second derivatives		Robust Standard Errors	
	Coefficient	Std.Error	t-value	Std.Error	t-value
Cst(M)	0.068847	0.020785	3.312	0.023222	2.965
Monday (M)	-0.028633	0.050560	-0.5663	0.055363	-0.5172
AR(1)	-0.059025	0.027620	-2.137	0.072989	-0.8087
Cst(V)	1.260808	0.042754	29.49	0.11311	11.15
Friday (V)	-0.364409	0.069532	-5.241	0.12689	-2.872
ARCH(Alpha1)	0.287649	0.030499	9.432	0.057891	4.969

Now we can compute the test for conditional heteroskedasticity, which shows that its absence is rejected:

```
Residual-Based Diagnostic for Conditional Heteroskedasticity of Tse (2002)
  RBD( 2) =  237.846    [0.0000000]
  RBD( 5) =  376.854    [0.0000000]
  RBD(10) =  987.233    [0.0000000]
```

4.5.2 Forecasts

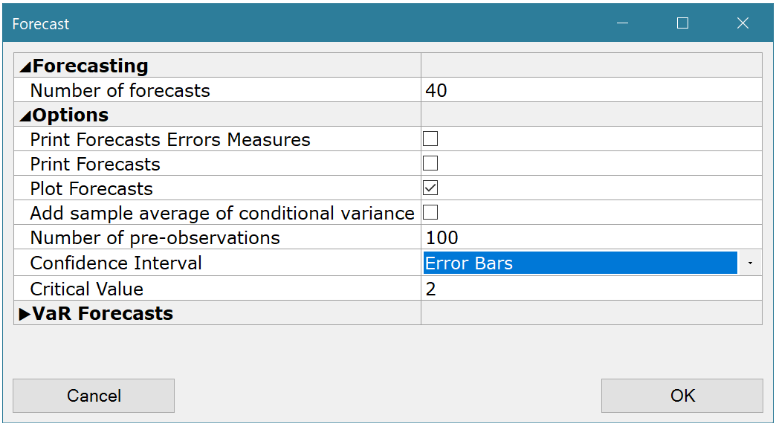
The main purpose of building and estimating a model with financial data is probably to produce forecasts. With the Forecast option, XlModeler also provides forecasting tools: forecasts of both the conditional mean and the conditional variance are available as well as several forecast error measures.

The first parameter to specify is the horizon h of the h-step-ahead forecasts. The default value is 10. Three options are available to:

1. print several forecasts error measures;
2. print the forecasts;
3. and make a graph of the forecasts.

Finally, graphical options are available for the standard error bands (error bands, bars or fans).

Our model has regressors in the mean and the variance, and these would have to be extended for out of sample forecasts (which is perfectly feasible, of course, as they are just calendar variables). Instead, we re-estimate the model, holding back 40 observations for forecasting. This allows us to make 40 out-of-sample (or dynamic) forecasts of the Nasdaq returns. Also set pre-observations to 100 and add error bands:



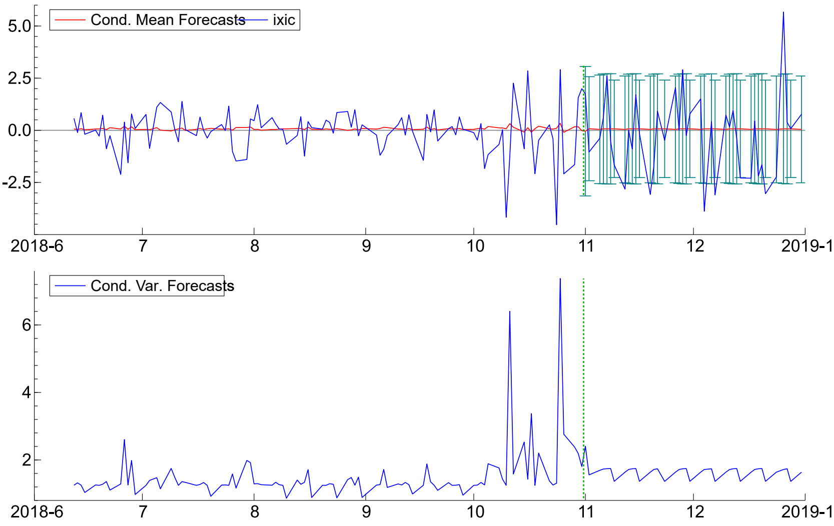


Figure 4.6 10-step-ahead forecasts of the Nasdaq

Our AR(1)-ARCH(1) model produces forecasts as shown in Figure 4.6. The forecast uncertainty bands are $\pm 2\hat{\sigma}_{t+h|t}$ which gives a 95 % confidence interval (note that the critical value 2 can be changed).

4.5.3 Exclusion Restrictions Dialog Box

The Exclusion Restrictions dialog box option allows you to select explanatory variables and test whether they are jointly significant. A more general form is the test for linear restrictions.

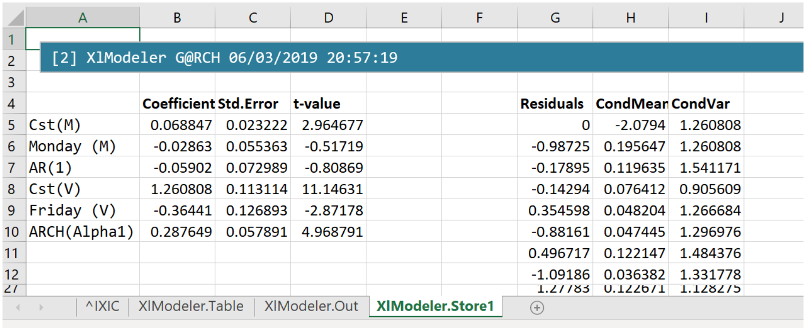
Mark all the variables you wish to include in the test in this Multiple-Selection List box. G@RCH tests whether the selected variables can be deleted from the model.

4.5.4 Linear Restrictions Dialog Box

Tests for linear restrictions are specified in the form of a matrix R, and a vector r. These are entered as one matrix $[R : r]$ in the dialog. This is more general than testing for exclusion restrictions.

4.5.5 Store Results

Finally, the coefficients, residuals, the conditional mean and the conditional variance can be stored in a separate worksheet:



	A	B	C	D	E	F	G	H	I	J
1										
2	[2] XlModeler G@RCH 06/03/2019 20:57:19									
3										
4		Coefficient	Std.Error	t-value			Residuals	CondMean	CondVar	
5	Cst(M)	0.068847	0.023222	2.964677			0	-2.0794	1.260808	
6	Monday (M)	-0.02863	0.055363	-0.51719			-0.98725	0.195647	1.260808	
7	AR(1)	-0.05902	0.072989	-0.80869			-0.17895	0.119635	1.541171	
8	Cst(V)	1.260808	0.113114	11.14631			-0.14294	0.076412	0.905609	
9	Friday (V)	-0.36441	0.126893	-2.87178			0.354598	0.048204	1.266684	
10	ARCH(Alpha1)	0.287649	0.057891	4.968791			-0.88161	0.047445	1.296976	
11							0.496717	0.122147	1.484376	
12							-1.09186	0.036382	1.331778	
27							1.27783	0.122671	1.128275	

^IXIC | XlModeler.Table | XlModeler.Out | XlModeler.Store1

4.5.6 Conclusions

While Engle (1982) is certainly the most important contribution in financial econometrics, the ARCH model is rarely used in practice due to its simplicity.

A useful generalization of ARCH is the GARCH model introduced by Bollerslev (1986). This is also a weighted average of past squared residuals, but it has declining weights that never go completely to zero. Even in its simplest form, as a GARCH(1,1), it has proven surprisingly successful in predicting conditional variances.

We will estimate a GARCH model next. After that, the many other variants that are available in XlModeler are introduced.

4.6 Other volatility models

4.6.1 GARCH Model

The GARCH(p, q) model of Bollerslev (1986) is:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2. \tag{4.10}$$

Alternatively, using lag polynomials $\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \ldots + \alpha_q L^q$ and $\beta(L)$:

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_t^2 + \beta(L)\sigma_t^2.$$

If all the roots of $|1 - \beta(L)| = 0$ lie outside the unit circle, we have:

$$\sigma_t^2 = \omega \left[1 - \beta(L)\right]^{-1} + \alpha(L)\left[1 - \beta(L)\right]^{-1} \varepsilon_t^2, \tag{4.11}$$

which may be seen as an ARCH(∞) process since the conditional variance linearly depends on all previous squared residuals. In that case, the conditional variance of y_t can become larger than the unconditional variance, given by

$$\sigma^2 \equiv E(\varepsilon_t^2) = \frac{\omega}{1 - \alpha(1) - \beta(1)},$$

if past realizations of ε_t^2 are larger than σ^2 (Palm, 1996).

Variance targeting for the GARCH model means replacing ω by $\sigma^2[1-\alpha(1)-\beta(1)]$.

Bollerslev (1986) has shown that for a GARCH$(1,1)$ with normal innovations, the kurtosis of y is $3[1-(\alpha_1+\beta_1)^2]/\left[1-(\alpha_1+\beta_1)^2-2\alpha_1^2\right] > 3$. The autocorrelations ρ_i of ε_t^2 have also been derived. For a stationary GARCH$(1,1)$, $\rho_1 = \alpha_1 + [\alpha_1^2\beta_1/(1-2\alpha_1\beta_1-\beta_1^2)]$, and $\rho_k = (\alpha_1+\beta_1)^{k-1}\rho_1$, $\forall k = 2,3,\ldots$ In other words, the autocorrelations decline exponentially with a decay factor of $\alpha_1+\beta_1$.

Estimation of the AR(1)-GARCH(1,1) model for ixic, with second derivatives for the variance, gives:

```
 Maximum Likelihood Estimation (Std.Errors based on Second derivatives)
                  Coefficient  Std.Error  t-value  t-prob
Cst(M)               0.075303   0.016758    4.493  0.0000
Monday (M)          -0.012476   0.040121  -0.3109  0.7559
AR(1)               -0.033023   0.018060   -1.828  0.0676
Cst(V)               0.042411   0.011859    3.576  0.0004
Friday (V)          -0.054172   0.051253   -1.057  0.2906
ARCH(Alpha1)         0.098302   0.010287    9.556  0.0000
GARCH(Beta1)         0.878398   0.012116    72.50  0.0000

No. Observations :      3522  No. Parameters  :         7
Mean (Y)         :   0.03197  Variance (Y)    :   1.65341
Skewness (Y)     :  -0.28344  Kurtosis (Y)    :  10.53423
Log Likelihood   : -5156.585  Alpha[1]+Beta[1]:   0.97650
```

The likelihood has improved from -5692.286 for the ARCH(1) model to -5156.585 for the current model. This is a very large change for adding just a single parameter. There is much less difference between robust and second-derivative t-values (you can check this easily).

We report below the same five misspecification tests as for the ARCH(1) model.

```
                   Statistic       t-Test      P-Value
Skewness            -0.51674       12.525   5.4592e-36
Excess Kurtosis       1.3589       16.473   5.7335e-61
Jarque-Bera           427.71         .NaN   1.3274e-93

Q-Statistics on Standardized Residuals
  --> P-values adjusted by 1 degree(s) of freedom
  Q(  5) =  2.93671   [0.5684726]
  Q( 10) =  13.8178   [0.1289565]
  Q( 20) =  28.1207   [0.0811332]
  Q( 50) =  61.8798   [0.1024133]

Q-Statistics on Squared Standardized Residuals
  --> P-values adjusted by 2 degree(s) of freedom
  Q(  5) =  3.54361   [0.3151500]
  Q( 10) =  17.2109   [0.0279860]*
  Q( 20) =  32.3848   [0.0197889]*
  Q( 50) =  53.7317   [0.2641310]

Adjusted Pearson Chi-square Goodness-of-fit test
# Cells(g)  Statistic      P-Value(g-1)     P-Value(g-k-1)
    40      172.0602         0.000000           0.000000
```

```
    50      186.2624          0.000000            0.000000
    60      201.9182          0.000000            0.000000
Rem.: k = 7 = # estimated parameters

Residual-Based Diagnostic for Conditional Heteroskedasticity of Tse (2002)
  RBD( 2) =   1.53567   [0.4640169]
  RBD( 5) =   3.58472   [0.6106088]
  RBD(10) =  17.2987    [0.0680111]
```

Unlike the ARCH(1) model, the Q-Statistics on standardized and squared standardized residuals, as well as the RBD test with various lag values suggest that the GARCH$(1,1)$ does a good job in modelling the dynamics of the first two conditional moments of the NASDAQ.

However, the adjusted Pearson Chi-square goodness-of-fit test (with different cell numbers) still points out some misspecification of the overall conditional distribution. The excess kurtosis is much reduced, but still significant.

Several authors have proposed to use a Student-t or GED distribution in combination with a GARCH model to model the fat tails of the high-frequency financial time-series. Furthermore, since the NASDAQ seems to be skewed, a skewed-Student distributions might be justified.

4.6.2 EGARCH Model

The Exponential GARCH (EGARCH) model, originally introduced by Nelson (1991), is re-expressed in Bollerslev and Mikkelsen (1996) as follows:

$$\log \sigma_t^2 = \omega + [1 - \beta(L)]^{-1} [1 + \alpha(L)] g(z_{t-1}).$$

The use of the $\log$ transformation of the conditional variance ensures that σ_t^2 is always positive.

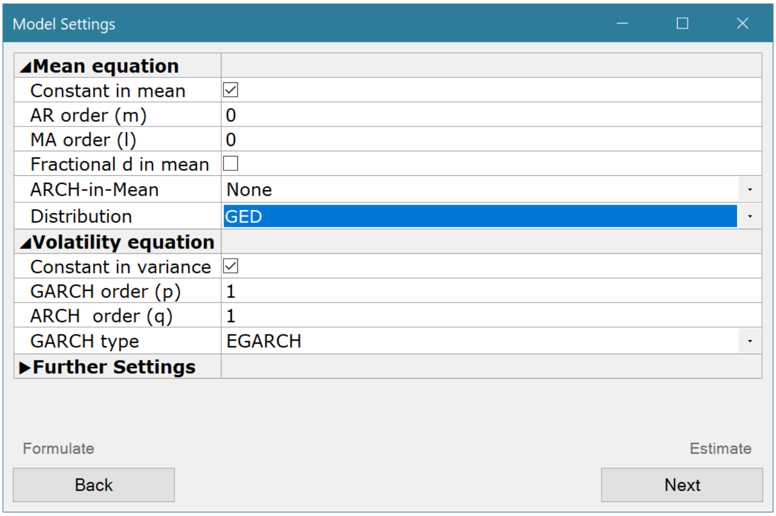
The value of $g(z_t)$ depends on several elements. Nelson (1991) notes that, *"to accommodate the asymmetric relation between stock returns and volatility changes (...) the value of $g(z_t)$ must be a function of both the magnitude and the sign of z_t"*.[1]:

$$g(z_t) \equiv \underbrace{\gamma_1 z_t}_{\text{sign effect}} + \underbrace{\gamma_2 [|z_t| - E|z_t|]}_{\text{magnitude effect}}$$

$E|z_t|$ depends on the assumption made on the unconditional density of z_t.

The output reported below corresponds to the ARMA(0,0)-EGARCH(1,1) with a GED distribution. In this case the model with AR(1) failed to converge, so we dropped that from the specification:

[1]Note that with the EGARCH parameterization of Bollerslev and Mikkelsen (1996), it is possible to estimate an EGARCH $(p, 0)$ since $\log \sigma_t^2$ depends on $g(z_{t-1})$, even when $q = 0$.



Interestingly, both θ_1 and θ_2 are significant. Note that the degree of freedom of the GED distribution is significantly lower than 2, confirming that the standardized residuals are fat-tailed.

```
 Maximum Likelihood Estimation (Std.Errors based on Second derivatives)
                  Coefficient  Std.Error  t-value  t-prob
Cst(M)               0.072928   0.015298    4.767  0.0000
Monday (M)          -0.021838   0.035502  -0.6151  0.5385
Cst(V)              -0.203273    0.13809   -1.472  0.1411
Friday (V)          -0.073880   0.067582   -1.093  0.2744
ARCH(Alpha1)        -0.201055    0.11557   -1.740  0.0820
GARCH(Beta1)         0.975918  0.0047214    206.7  0.0000
EGARCH(Theta1)      -0.193468   0.028299   -6.837  0.0000
EGARCH(Theta2)       0.141604   0.021003    6.742  0.0000
G.E.D.(DF)           1.370211   0.045900    29.85  0.0000

No. Observations :       3522  No. Parameters  :         9
Mean (Y)         :    0.03197  Variance (Y)    :   1.65341
Skewness (Y)     :   -0.28344  Kurtosis (Y)    :  10.53423
Log Likelihood   :  -5026.338
```

4.6.3 GJR Model

This popular model is proposed by Glosten, Jagannathan, and Runkle (1993):

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} (\alpha_i \varepsilon_{t-i}^2 + \gamma_i S_{t-i}^- \varepsilon_{t-i}^2) + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2,$$

where S_t^- is a dummy variable that takes the value 1 when ε_t is negative and 0 otherwise.

In this model, the impact of ε_t^2 on the conditional variance σ_t^2 is different when ε_t is positive or negative. An attractive feature of the GJR model is that the null hypothesis of no leverage effect is easy to test: $\gamma_1 = \ldots = \gamma_q = 0$ implies that the news impact curve is symmetric, i.e. past positive shocks have the same impact on today's volatility as past negative shocks.

The output reported below suggests the presence of such an effect on the NASDAQ since $\hat{\gamma}_1 = 0.2$ with a robust t-value of 7.1:

```
 Robust Standard Errors (Sandwich formula)
                  Coefficient  Std.Error  t-value  t-prob
Cst(M)               0.035811   0.017387    2.060  0.0395
Monday (M)          -0.005590   0.040008  -0.1397  0.8889
AR(1)               -0.027368   0.017660   -1.550  0.1213
Cst(V)               0.038964   0.014363    2.713  0.0067
Friday (V)          -0.010132   0.068176  -0.1486  0.8819
ARCH(Alpha1)        -0.018158  0.0081774   -2.220  0.0264
GARCH(Beta1)         0.886526   0.013427    66.02  0.0000
GJR(Gamma1)          0.198282   0.027992    7.083  0.0000

No. Observations :      3522  No. Parameters   :         8
Mean (Y)         :   0.03197  Variance (Y)     :   1.65341
Skewness (Y)     :  -0.28344  Kurtosis (Y)     :  10.53423
Log Likelihood   : -5085.808
```

4.6.4 APARCH Model

The APARCH (p, q) has been introduced by Ding, Granger, and Engle (1993):

$$\sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i \left(|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i}\right)^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta,$$

where $\delta > 0$ and $-1 < \gamma_i < 1$ $(i = 1, ..., q)$. The parameter δ plays the role of a Box-Cox transformation of σ_t while γ_i reflects the so-called leverage effect. Properties of the APARCH model are studied in He and Teräsvirta (1999a), He and Teräsvirta (1999b).

The APARCH includes seven other ARCH extensions as special cases:

- ARCH when $\delta = 2$, $\gamma_i = 0$ $(i = 1, \ldots, p)$ and $\beta_j = 0$ $(j = 1, \ldots, p)$.
- GARCH when $\delta = 2$ and $\gamma_i = 0$ $(i = 1, \ldots, p)$.
- Taylor (1986)/Schwert (1990)'s GARCH when $\delta = 1$, and $\gamma_i = 0$ $(i = 1, \ldots, p)$.
- GJR when $\delta = 2$.
- TARCH, Zakoian (1994), when $\delta = 1$.
- NARCH, Higgins and Bera (1992), when $\gamma_i = 0$ $(i = 1, \ldots, p)$ and $\beta_j = 0$ $(j = 1, \ldots, p)$.
- Log-ARCH of Geweke (1986) and Pentula (1986), when $\delta \to 0$.

4.6.5 IGARCH Model

In many high-frequency time-series applications, the conditional variance estimated using a GARCH$(1, 1)$ process exhibits a strong persistence, that is $\alpha_1 + \beta_1 \approx 1$. The IGARCH$(p, q)$ model sets $\alpha(1) + \beta(1) = 1$.

4.6.6 RiskMetrics^{TM}

In October 1994, the risk management group at J.P. Morgan released a technical document describing its internal market risk management methodology (J.P.Morgan, 1996). This methodology, called RiskMetrics^{TM} soon became a standard in the market risk measurement due to its simplicity.

The basic RiskMetrics^{TM} model is an IGARCH(1,1) model where the ARCH and GARCH coefficients are fixed:

$$\sigma_t^2 = \omega + (1-\lambda)\varepsilon_{t-1}^2 + \lambda\sigma_{t-1}^2, \tag{4.12}$$

where $\omega = 0$ and λ is generally set to 0.94 with daily data and to 0.97 with weekly data.

To illustrate the need for flexible ARCH-type models, here is the output of the Box-Pierce test on squared standardized residuals and the RBD test applied after the estimation of the RiskMetrics model (including an AR(1) term and the two dummy variables). Furthermore. the likelihood ratio test statistic is about 57 for three restrictions. There is no doubt that the RiskMetrics specification is not appropriate.

```
 Maximum Likelihood Estimation (Std.Errors based on Second derivatives)
                  Coefficient  Std.Error  t-value  t-prob
Cst(M)               0.064194   0.017017    3.772  0.0002
Monday (M)          -0.009739   0.041319  -0.2357  0.8137
AR(1)               -0.032809   0.017562   -1.868  0.0618
Friday (V)           0.034130  0.0053469    6.383  0.0000
ARCH(Alpha1)         0.060000
GARCH(Beta1)         0.940000

No. Observations :       3522  No. Parameters   :         4
Mean (Y)         :    0.03197  Variance (Y)     :   1.65341
Skewness (Y)     :   -0.28344  Kurtosis (Y)     :  10.53423
Log Likelihood   : -5184.951

Q-Statistics on Squared Standardized Residuals
  --> P-values adjusted by 2 degree(s) of freedom
  Q(  5) =  25.1700   [0.0000142]**
  Q( 10) =  42.9937   [0.0000009]**
  Q( 20) =  60.3281   [0.0000018]**
  Q( 50) =  86.4578   [0.0005570]**

Adjusted Pearson Chi-square Goodness-of-fit test
# Cells(g)  Statistic      P-Value(g-1)     P-Value(g-k-1)
    40      179.9648         0.000000          0.000000
    50      206.7053         0.000000          0.000000
    60      223.2470         0.000000          0.000000

Residual-Based Diagnostic for Conditional Heteroskedasticity of Tse (2002)
  RBD( 2) =  5.84439   [0.0538155]
  RBD( 5) =  9.52249   [0.0899528]
  RBD(10) =  21.6940   [0.0167415]
```

4.6.7 Fractionally Integrated Models

Volatility tends to change quite slowly over time, and, as shown in Ding, Granger, and Engle (1993) among others, the effects of a shock can take a considerable time to decay: in their study of the daily S&P500 index, they find that the squared returns series has positive autocorrelations over more than 10 years. Therefore the distinction between stationary and unit root processes seems to be far too restrictive.

To mimic the behavior of the correlogram of the observed volatility, Baillie, Bollerslev, and Mikkelsen (1996) (hereafter denoted BBM) introduce the Fractionally Integrated GARCH (FIGARCH) model. The conditional variance of the FIGARCH (p, d, q) is given by:

$$\sigma_t^2 = \underbrace{\omega[1-\beta(L)]^{-1}}_{\omega^*} + \underbrace{\left\{1-[1-\beta(L)]^{-1}\phi(L)(1-L)^d\right\}}_{\lambda(L)}\varepsilon_t^2,$$

or $\sigma_t^2 = \omega^* + \sum_{i=1}^{\infty}\lambda_i L^i \varepsilon_t^2 = \omega^* + \lambda(L)\varepsilon_t^2$, with $0 \le d \le 1$.

When estimating a FIGARCH $(1, d, 1)$ model by QML on the NASDAQ dataset (noticeably slower) we obtain a loglikelihood of -5158.790 in contrast to a GARCH(1,1) value of -5156.585. So in this case the contribution of the fractional integration is fairly small.

4.6.8 Spline-GARCH Model

Unlike most existing GARCH-type models, the Spline-GARCH model of Engle and Rangel (2008) does not assume that the unconditional variance (when it exists) is constant over time but allows it to change smoothly as a function of time.

Now (4.3) is extended by including a factor τ_t as follows:

$$\varepsilon_t = \tau_t s_t z_t. \tag{4.13}$$

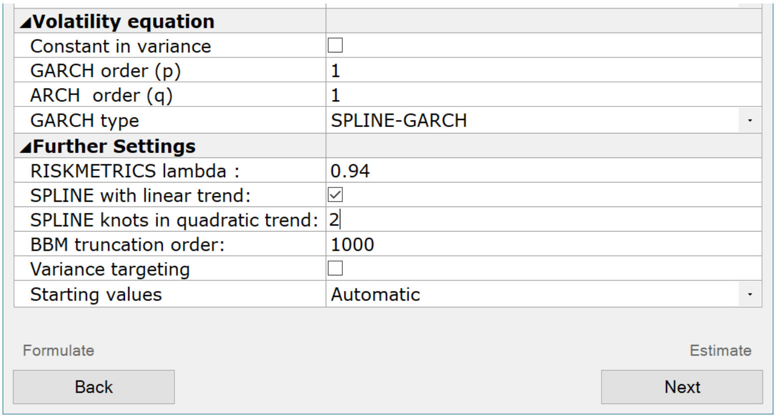
The factor τ_t is an exponential quadratic spline function with k knots and is multiplied by a GARCH(p, q) component. Using a GARCH(1,1):

$$s_t^2 = 1-(\alpha_1+\beta_1)+\alpha_1(\varepsilon_{t-1}/\tau_{t-1})^2+\beta_1 s_{t-1}^2, \tag{4.14}$$

$$\tau_t^2 = \omega\exp\left(\delta_0 t + \sum_{i=1}^{k}\delta_i[(t-t_{i-1})_+]^2\right), \tag{4.15}$$

where ω, α_1, β_1 and δ_i for $i = 0, 1, \ldots, k$ are parameters, $x_+ = x$ if $x > 0$ and 0 otherwise, and $\{t_0 = 0, t_1, \ldots, t_{k-1}\}$ are time indices partitioning the time span into k equally spaced intervals. The specification of s_t^2 may be chosen among other available GARCH equations provided that $E(s_t) = 1$ (which implies an identification constraint for the intercept).

To illustrate, we estimate the AR(1)-Spline-GARCH$(1, 1)$ on the returns of the Nasdaq index.



The number of knots is set to 2 (this can be selected by minimixing the SC information criterion).

```
 Robust Standard Errors (Sandwich formula)
                  Coefficient  Std.Error  t-value  t-prob
Cst(M)               0.074977   0.017115    4.381  0.0000
Monday (M)          -0.012732   0.039268  -0.3242  0.7458
AR(1)               -0.032042   0.017620   -1.818  0.0691
Cst(V)               0.672532    0.13954    4.820  0.0000
Friday (V)          -0.058664   0.074794  -0.7843  0.4329
Spline_0 (V)         7.159803     1.7730    4.038  0.0001
Spline_1 (V)       -11.131828     2.9067   -3.830  0.0001
Spline_2 (V)        18.966135     6.0312    3.145  0.0017
ARCH(Alpha1)         0.099659   0.012779    7.799  0.0000
GARCH(Beta1)         0.866780   0.015749    55.04  0.0000

No. Observations :       3522  No. Parameters   :        10
Mean (Y)         :    0.03197  Variance (Y)     :   1.65341
Skewness (Y)     :   -0.28344  Kurtosis (Y)     :  10.53423
Log Likelihood   :  -5147.146  Alpha[1]+Beta[1]:   0.96644
```

Figure 4.7 displays the daily returns of the Nasdaq, the estimated spline component $\hat{\tau}_t^2$, which clearly reflects the increase of volatility at the end of the sample, and the conditional variance $\hat{\sigma}_t^2$.

4.6.9 Generalized Autoregressive Score (GAS) Models

It is well known that financial series occasionally exhibit large changes, also known as jumps. The impact of jumps has been modelled assuming a Poisson or a Bernoulli jump distribution which, when combined with a normal distribution for the Brownian motion part, leads to Poisson or Bernoulli mixtures of distributions for financial returns. Alternatively some studies assume fat tail distributions such as the (skewed) student-t or the generalized error distribution to account for the occurrence of large changes in returns. Several authors have shown that these jumps affect future volatility less than what standard volatility models would predict. Many volatility models, such as GARCH, are based on the assumption that each return observation has the same relative impact on

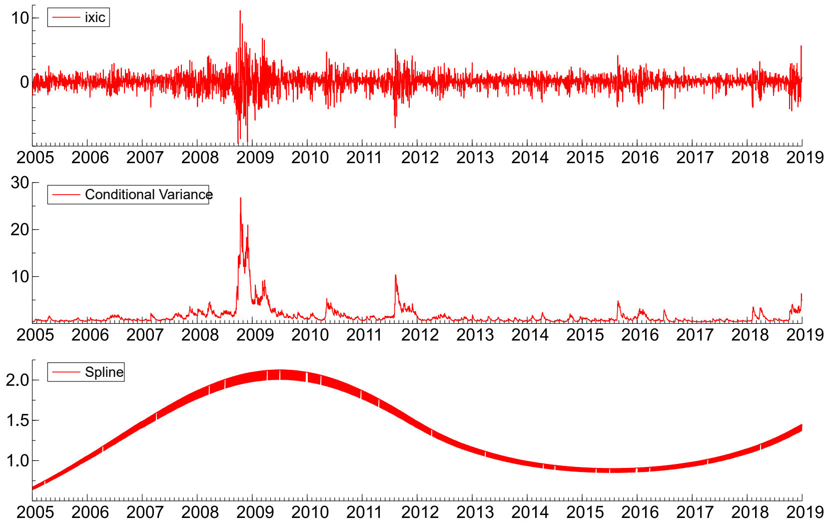


Figure 4.7 Daily returns of the Nasdaq, conditional variance, and spline component for a spline-GARCH with two knots

future volatility, regardless of the magnitude of the return. This assumption is at odds with an increasing body of evidence indicating that the largest return observations have a relatively smaller effect on future volatility than smaller shocks.

To overcome this problem, Harvey and Chakravarty (2008) and Creal, Koopman, and Lucas (2012) independently proposed a novel way to deal with large returns in a GARCH context. Their models rely on a potentially non-normal distribution for the innovations z_t in (4.3) and a GARCH-type equation for the conditional variance derived from the conditional score of the assumed distribution with respect to the second moment.

Start by rewriting the GARCH(1,1) model as:

$$\sigma_t^2 = \omega + \alpha_1 \underbrace{z_{t-1}^2 \sigma_{t-1}^2}_{\varepsilon_{t-1}^2} + \beta_1 \sigma_{t-1}^2$$

or equivalently

$$\sigma_t^2 = \omega + \alpha_1 \underbrace{(z_{t-1}^2 - 1)}_{u_{t-1}} \sigma_{t-1}^2 + \underbrace{(\alpha_1 + \beta_1)}_{\phi_1} \sigma_{t-1}^2. \tag{4.16}$$

The specification of the GAS(1,1) model of Harvey and Chakravarty (2008) combined with a normal, Student-t, GED or Skewed-Student distribution defines u_t in (4.16):

$$u_t = z_t^2 - 1 \text{ if } z_t \sim N(0,1); \tag{4.17}$$

$$u_t = \frac{(\nu+1)z_t^2}{\nu - 2 + z_t^2} - 1 \text{ if } z_t \sim t(0,1,\nu); \tag{4.18}$$

$$u_t = 0.5\nu |z_t|^{\nu}/\lambda_{\nu}^{\nu} - 1 \text{ if } z_t \sim GED(0,1,\nu); \tag{4.19}$$

$$u_t = \frac{(\nu+1)z_t z_t^*}{(\nu-2)g_t \xi^{I_t}} - 1 \text{ if } z_t \sim SKST(0,1,\xi,\nu). \tag{4.20}$$

Harvey and Chakravarty (2008) call the above GAS model with a T distribution 'Beta-t-GARCH' because, for this distribution, $(u_t+1)/(\nu+1)$ has a Beta distribution. Note that the normal GARCH(1,1) is identical to the normal GAS(1,1) model.

As example consider the stock price of Bristol-Myers Squibb (BMY). Daily returns in % of BMY (on the period 1999-2008) are plotted in the top panel of Figure 4.8. Bristol-Myers Squibb, one of the largest Pharmaceutical companies in the US, withdrew a New Drug Application for a drug called Omapatrilat on April 19, 2000. This was generally seen as a huge blow to the company as it was meant to be the company's next blockbuster. The product was expected to be a topseller amongst all pharmaceuticals. The market reacted heavily with a 30% loss on a single day. This was a once-off event which influenced future profitability, but not the further functioning of the company. The market immediately adapted to the new information and the stock's returns remained calm afterwards.

The bottom panel of Figure 4.8 plots the estimated conditional standard deviation of BMY. The dotted line corresponds to $\hat{\sigma}_t$ obtained for a GARCH(1,1) with a skewed-Student distribution, whereas the solid blue line is for the equivaleny GAS(1,1) model. The difference in response to the large jump is striking.

The data set is supplied as BMY.xlsx, so you should be able to replicate the following estimates:

```
 Robust Standard Errors (Sandwich formula)
                ---------- GAS(1,1) ---------  ---------- GARCH(1.1) ------
                Coefficient  Std.Er  t-value  Coefficient  Std.Er  t-value
Cst(M)            -0.017871   0.027 -0.656     -0.015598   0.027   -0.572
Cst(V)             0.018856   0.010  1.916      0.021660   0.011    1.977
GAS(Alpha1)        0.094193   0.020  4.683     -
GAS(Phi1)          0.996115   0.005  215.4     -
ARCH(Alpha1)     -                              0.073753   0.020    3.643
GARCH(Beta1)     -                              0.922380   0.021   43.46
Asymmetry          0.011241   0.028  0.398      0.016231   0.028    0.583
Tail               7.458755   1.136  6.565      7.206739   1.118    6.444

No. Observations :       2489             No. Observations :       2489
Log Likelihood   : -4602.530              Log Likelihood   : -4617.036
```

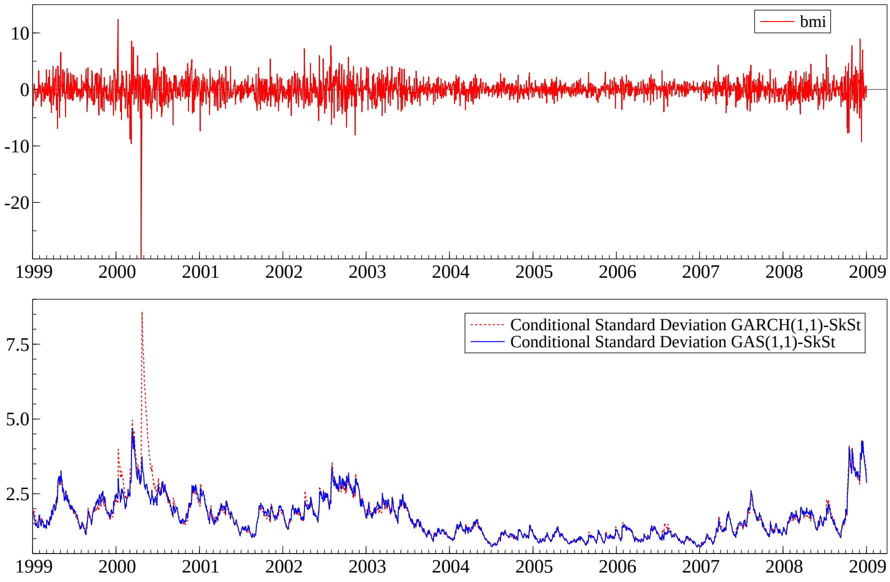


Figure 4.8 Daily returns in % of Bristol-Myers Squibb (BMY) and estimated conditional standard deviation

References

Baillie, R. T., T. Bollerslev, and H. O. Mikkelsen (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 74*, 3–30.

Banerjee, A., J. J. Dolado, J. W. Galbraith, and D. F. Hendry (1993). *Co-integration, Error Correction and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press.

Banerjee, A., J. J. Dolado, and R. Mestre (1998). Error-correction mechanism tests for cointegration in a single equation framework. *Journal of Time Series Analysis 19*, 267–283.

Blasques, F., K. Lasak, S. Koopman, and A. Lucas (2016). In-sample confidence bands and out-of-sample forecast bands for time-varying parameters in observation driven models. *International Journal of Forecasting 32*, 875–887.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*, 307–327.

Bollerslev, T. and H. O. Mikkelsen (1996). Modeling and pricing long-memory in stock market volatility. *Journal of Econometrics 73*, 151–184.

Bollerslev, T. and J. M. Wooldridge (1992). Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews 11*, 143–172.

Clements, M. P. and D. F. Hendry (1998). *Forecasting Economic Time Series*. Cambridge: Cambridge University Press.

Creal, D., S. Koopman, and A. Lucas (2012). A general framework for observation driven time-varying parameter models. Forthcoming in Journal of Applied Econometrics.

Davidson, J. E. H., D. F. Hendry, F. Srba, and J. S. Yeo (1978). Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom. *Economic Journal 88*, 661–692. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000; and in Campos, J., Ericsson, N.R. and Hendry, D.F. (eds.), *General to Specific Modelling*. Edward Elgar, 2005.

Ding, Z., C. W. J. Granger, and R. F. Engle (1993). A long memory property of stock market returns and a new model. *Journal of Empirical Finance 1*, 83–106.

Doornik, J. A. (2008). Encompassing and automatic model selection. *Oxford Bulletin of Economics and Statistics 70*, 915–925.

Doornik, J. A. (2009). Autometrics. In J. L. Castle and N. Shephard (Eds.), *The Methodology and Practice of Econometrics: Festschrift in Honour of David F. Hendry*. Oxford: Oxford University Press.

Engle, R. and J. Rangel (2008). The spline-GARCH model for low-frequency volatility and its global macroeconomic causes. *Review of Financial Studies 21*, 1187–1222.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica 50*, 987–1007.

Engle, R. F. and G. González-Rivera (1991). Semiparametric arch model. *Journal of Business and Economic Statistics 9*, 345–360.

Geweke, J. (1986). Modeling the persistence of conditional variances: A comment. *Econometric Reviews 5*, 57–61.

Glosten, L. R., R. Jagannathan, and D. E. Runkle (1993). On the relation between expected value and the volatility of the nominal excess return on stocks. *Journal of Finance 48*, 1779–1801.

Harvey, A. and T. Chakravarty (2008). Beta-t-(e)garch. Working Paper series, University of Cambridge.

He, C. and T. Teräsvirta (1999a). Higher-order dependence in the general power arch process and a special case. Stockholm School of Economics, Working Paper Series in Economics and Finance, No. 315.

He, C. and T. Teräsvirta (1999b). Statistical properties of the asymmetric power arch process. In R. Engle and H. White (Eds.), *Cointegration, causality, and forecasting. Festschrift in honour of Clive W.J. Granger*, Chapter 19, pp. 462–474. Oxford: Oxford University Press.

Hendry, D. F. (1987). Econometric methodology: A personal perspective. In T. F. Bewley (Ed.), *Advances in Econometrics*, pp. 29–48. Cambridge: Cambridge University Press. Reprinted in Campos, J., Ericsson, N.R. and Hendry, D.F. (eds.), *General to Specific Modelling*. Edward Elgar, 2005.

Hendry, D. F. and J. A. Doornik (2013). *Empirical Econometric Modelling using PcGive: Volume I* (7th ed.). London: Timberlake Consultants Press.

Hendry, D. F. and J. A. Doornik (2014). *Empirical Model Discovery and Theory Evaluation: Automatic Selection Methods in Econometrics*. Cambridge, MA: MIT Press.

Hendry, D. F. and H.-M. Krolzig (1999). Improving on 'Data mining reconsidered' by K.D. Hoover and S.J. Perez. *Econometrics Journal 2*, 202–219. Reprinted in J. Campos, N.R. Ericsson and D.F. Hendry (eds.), *General to Specific Modelling*. Edward Elgar, 2005.

Hendry, D. F. and G. E. Mizon (1978). Serial correlation as a convenient simplification, not a nuisance: A comment on a study of the demand for money by the Bank of England. *Economic Journal 88*, 549–563. Reprinted in Hendry, D. F., *Econometrics: Alchemy or Science?* Oxford: Blackwell Publishers, 1993, and Oxford University Press, 2000; and in Campos, J., Ericsson, N.R. and Hendry, D.F. (eds.), *General to Specific Modelling*. Edward Elgar, 2005.

Higgins, M. L. and A. K. Bera (1992). A class of nonlinear arch models. *International Economic Review 33*, 137–158.

Hoover, K. D. and S. J. Perez (1999). Data mining reconsidered: Encompassing and the general-to-specific approach to specification search. *Econometrics Journal 2*, 167–191.

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

J.P.Morgan (1996). *Riskmetrics Technical Document, 4th ed.* New York: J.P.Morgan.

Lambert, P. and S. Laurent (2000). Modelling skewness dynamics in series of financial data. Discussion Paper, Institut de Statistique, Louvain-la-Neuve.

Makridakis, S., S. C. Wheelwright, and R. C. Hyndman (1998). *Forecasting: Methods and Applications* (3rd ed.). New York: John Wiley and Sons.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica 59*, 349–370.

Nocedal, J. and S. J. Wright (1999). *Numerical Optimization*. New York: Springer-Verlag.

Palm, F. C. (1996). GARCH models of volatility. In G. Maddala and C. Rao (Eds.), *Handbook of Statistics*, pp. 209–240. Amsterdam: Elsevier Science.

Pentula, S. G. (1986). Modeling the persistence of conditional variances: A comment. *Econometric Reviews 5*, 71–74.

Schwert, W. (1990). Stock volatility and the crash of '87. *Review of financial Studies 3*, 77–102.

Taylor, S. J. (1986). *Modelling Financial Time Series*. Chichester: John Wiley.

Weiss, A. A. (1986). Asymptotic theory for arch models: Estimation and testing. *Econometric Theory 2*, 107–131.

Zakoian, J.-M. (1994). Threshold heteroskedasticity models. *Journal of Economic Dynamics and Control 15*, 931–955.