

Chapter 2

Regression Model

2.1 Introduction

This chapter starts with the introduction to a linear regression analysis, estimation and inference methods. Regression analysis is widely used tool in financial econometrics. They are used to describe and evaluate the relationship between financial variables, perform forecasting tasks.

This chapter provides only a short and brief description of main tools used in the regression analysis. More detailed discussion and deeper theoretical background can be found in Greene (2000), Hamilton (1994), Hayashi (2000), Verbeek (2008), Mills (1999), Zivot and Wang (2006).

2.2 Linear Regression Model

Consider the linear regression model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \dots + \beta_k X_{ki} + u_i = \mathbf{X}'_i \beta + u_i, \quad i = 1, \dots, n, \quad (2.2.1)$$

where $\mathbf{X}_i = [1, X_{2i}, \dots, X_{ki}]'$ is a $k \times 1$ vector of explanatory variables, $\beta = (\beta_1, \dots, \beta_k)'$ is a $k \times 1$ vector of coefficients, and u_i is a random error term. In matrix form the model is expressed as

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{u}, \quad (2.2.2)$$

where \mathbf{Y} and β are $n \times 1$ vectors and \mathbf{X} is a $n \times k$ matrix.

The standard assumptions of the linear regression model are:

1. the linear model (2.2.2) is correctly specified;
2. the regressors \mathbf{X}_i are uncorrelated with the error term \mathbf{u} : $E[\mathbf{X}_i u_i] = 0$ for all $i = 1, \dots, n$;

3. $E[\mathbf{X}_i\mathbf{X}_i'] = \sigma_{XX}$ is of full rank k ;
4. u_i are independently identically distributed (iid) with mean zero and constant variance σ^2 .

Ordinary Least Squares (OLS) estimation is based on minimizing the residual sum of squares RSS . The fitted model is

$$Y_i = \mathbf{X}_i'\hat{\beta} + \hat{u}_i, \quad i = 1, \dots, n,$$

where

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

and $\hat{u}_i = \mathbf{Y}_i - \hat{\mathbf{Y}}_i = Y_i - \mathbf{X}_i'\hat{\beta}$. An unbiased estimator of the regression variance is $\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}'\hat{\mathbf{u}}}{n-k}$.

Under the assumptions described above, the OLS estimates $\hat{\beta}$ are consistent and asymptotically normally distributed. A consistent estimator of the asymptotic variance of the parameters estimator is

$$\text{var}[\hat{\beta}] = \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}. \quad (2.2.3)$$

Estimated standard errors $se[\hat{\beta}_i]$ for individual parameter estimators $\hat{\beta}_i$ are given by the square root of the diagonal elements of (2.2.3).

Goodness of fit is summarized by the R^2 of the regression $R^2 = 1 - \frac{RSS}{TSS}$, where $TSS = \sum_{i=1}^n (Y_i - \bar{Y})^2$. The coefficient R^2 measures the percentage of the variation of the dependent variable \mathbf{Y} that is explained by the variation of the regressors \mathbf{X} . The usual R^2 has the undesirable feature of never decreasing as more variables are added to the regression, even if the extra variables are irrelevant. A common way to solve the problem is to adjust R^2 for degrees of freedom; this gives $\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-k}$. The adjusted \bar{R}^2 may decrease with the addition of variables with low explanatory power.

2.2.1 Hypothesis testing

Suppose that we need to test the null hypothesis

$$H_0: \beta_j = \beta_j^0.$$

The OLS test statistic for testing this hypothesis (also called t-statistic) is

$$t = \frac{\hat{\beta}_j - \beta_j^0}{se[\hat{\beta}_j]},$$

which is asymptotically distributed $N(0, 1)$ under the null hypothesis. With the additional assumption of iid Gaussian error term, $\hat{\beta}_j$ is normally distributed and the t-statistic follows Student's t distribution with $n - k$ degrees of freedom.

More general linear restrictions hypotheses of the form $H_0: R\beta = r$, where R is a fixed $m \times k$ matrix of rank m and r is a $m \times 1$ vector, are tested using the Wald statistic

$$F = \frac{(R\hat{\beta} - r)' (R(\mathbf{X}'\mathbf{X})^{-1} R') (R\hat{\beta} - r)}{\text{var}[\hat{\sigma}^2]}.$$

Under the null, the Wald statistic is asymptotically distributed χ_m^2 . Under the Gaussian assumptions of residuals, $F/m \sim F_{m, n-k}$.

The statistical significance of all of the regressors excluding the intercept is captured by the F-statistic

$$F = \frac{ESS/(k - 1)}{RSS/n - k} \frac{R^2/k}{(1 - R^2)/(n - k)},$$

which is distributed $F_{k-1, n-k}$ under the null hypothesis that all slope coefficients are zero.

Need help with your dissertation?

Get in-depth feedback & advice from experts in your topic area. Find out what you can do to improve the quality of your dissertation!

Get Help Now



Go to www.helpmyassignment.co.uk for more info

 Helpmyassignment



2.2.2 Residual diagnostics

If the classical assumptions of the linear regression model do not hold it may lead to inconsistent, inefficient estimates. There are several residual diagnostic statistics are usually reported along with the regression results to check the validity of the model predictions.

Two common problems with the regression assumptions are heteroscedasticity and autocorrelation of the error terms. Heteroscedasticity means that variances of error terms are not constant from observation to observation. Autocorrelation means presence of series correlation between error terms. In both cases, the OLS estimator is unbiased and consistent but not efficient anymore. Moreover, the standard formula for computing the variance of the parameters estimators (2.2.3) is not valid anymore which may lead to wrong conclusions. If the variance-covariance matrix of error terms $\text{var}[\mathbf{u}] = \sigma^2\Omega$ then

$$\text{var}[\hat{\beta}] = \hat{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\Omega\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}.$$

One way of obtaining an efficient estimate of the regression parameters is to use Generalized Least Squares (GLS) method. The GLS estimator is given by $\hat{\beta}_{GLS} = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{Y}$ with variance $\text{var}[\hat{\beta}] = \hat{\sigma}^2 (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}$.

If the matrix Ω is not known, one can use White's heteroscedasticity consistent estimator of standard errors of the OLS estimators. The matrix

$$\hat{\text{var}}[\hat{\beta}] = \left(\sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i' \right)^{-1} \left(\sum_{i=1}^n \hat{u}_i^2 \mathbf{X}_i\mathbf{X}_i' \right) \left(\sum_{i=1}^n \mathbf{X}_i\mathbf{X}_i' \right)^{-1}$$

can be used as an estimate of the true variance of the OLS estimator.

There are several testing procedures to detect heteroscedasticity. The White test suggest to estimate an auxiliary regression of the squared OLS residuals on a constant and all regressors, their squares and cross products. Under the null hypothesis of homoscedasticity nR^2 statistic is asymptotically distributed $\chi^2(q)$, where q is the number of variables in the auxiliary regression minus one. If the value of statistic is large, the null hypothesis of homoscedasticity in residuals is rejected.

Another test for heteroscedasticity is Breusch-Godfrey-Pagan test. It suggests to regress squared residuals from the initial regression scaled by $\hat{\sigma}^2 = \sum \hat{u}_i^2/n$ on a set of known variables \mathbf{Z}_t (they also could be regressors but not restricted to). Under the null of homoscedasticity, the scaled $\frac{1}{2}ESS$ from the auxiliary regression follows asymptotically $\chi^2(p-1)$ distribution, where p is a number of auxiliary variables \mathbf{Z}_t .

The most common diagnostic statistics for presence of autocorrelation based on the estimated residuals $\hat{\mathbf{u}}_i$ is the Durbin-Watson statistic DW . It is defined as

$$DW = \frac{\sum_{i=2}^n (\hat{u}_i - \hat{u}_{i-1})^2}{\sum_{i=1}^n \hat{u}_i^2} \quad (2.2.4)$$

For large n the Durbin-Watson statistics can be approximated $DW = 2(1 - \hat{\rho})$, where $\hat{\rho}$ is the estimated correlation between \hat{u}_i and \hat{u}_{i-1} . Thus, the range of values of DW is from 0 to 4. Values of DW around 2 indicate no serial correlation in the error terms, values less than 2 suggest positive serial correlation, and values greater than 2 suggest negative serial correlation.

Exact critical values for a general case cannot be tabulated; however, Durbin and Watson (1950) established upper and lower bounds (d_U and d_L respectively) for the critical values. The testing procedure is as follows:

- if $DW < d_L$ we reject the null hypothesis of no autocorrelation in favour of positive first-order autocorrelation;
- if $DW > d_U$ we do not reject the null hypothesis

The bounds for critical values in the case of negative autocorrelation alternative are $4 - d_U$ and $4 - d_L$. The values of the bounds can be found in Savin and White (1977); some of the are tabulated in Table 2.1.

Table 2.1: Lower and Upper bounds for 5% critical values of the Durbin-Watson test

n	Number of regressors							
	$k = 3$		$k = 5$		$k = 7$		$k = 9$	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
25	1.206	1.550	1.038	1.767	0.868	2.012	0.702	2.280
50	1.462	1.628	1.378	1.721	1.291	1.822	1.201	1.930
75	1.571	1.680	1.515	1.739	1.458	1.801	1.399	1.867
100	1.634	1.715	1.592	1.758	1.550	1.803	1.506	1.850
200	1.748	1.789	1.728	1.810	1.707	1.831	1.686	1.852

The Breusch-Godfrey test for autocorrelation considers the regression of the OLS residuals \hat{u}_i upon its lag \hat{u}_{i-1} . This auxiliary regression produces an estimate for the first-order autocorrelation coefficient $\hat{\rho}$ and provides a standard error to this

estimate. In general case the test is easily extended to higher orders of autocorrelation by including additional lags of the residual. Testing the null hypothesis of no autocorrelation is equivalent to testing the significance of the auxiliary regression.

Another common diagnostic for serial correlation is the Ljung-Box modified Q statistic. The Q-statistic at lag q is a test statistic for the null hypothesis of no autocorrelation up to order q and is computed as:

$$Q = n(n+2) \sum_{j=1}^q \frac{\hat{\rho}_j^2}{n-j} \sim \chi_q^2,$$

where $\hat{\rho}_j$ is the j -th autocorrelation.

The most often used diagnostic statistic to test for normality of the residuals is the Jarque-Bera test statistics. It measures the difference of the skewness and kurtosis of the series with those from the normal distribution. The statistic is computed as:

$$JB = \frac{n}{6} \left(S^2 + \frac{(K-3)^2}{4} \right) \sim \chi_2^2,$$

where S is the skewness, and K is the kurtosis. We reject the null hypothesis of normality if a Jarque-Bera statistic exceeds the corresponding critical value.



Brain power

By 2020, wind could provide one-tenth of our planet's electricity needs. Already today, SKF's innovative know-how is crucial to running a large proportion of the world's wind turbines.

Up to 25 % of the generating costs relate to maintenance. These can be reduced dramatically thanks to our systems for on-line condition monitoring and automatic lubrication. We help make it more economical to create cleaner, cheaper energy out of thin air.

By sharing our experience, expertise, and creativity, industries can boost performance beyond expectations. Therefore we need the best employees who can meet this challenge!

The Power of Knowledge Engineering

Plug into The Power of Knowledge Engineering.
Visit us at www.skf.com/knowledge

SKF



2.2.3 Example: Factor Model

Fama and French (1993) suggested a three-factor model to explain the expected stock return premium required by investors. The three factors are

- The excess return of the market portfolio ($R_{mt} - r_{ft}$);
- The difference between the expected returns on portfolios of small and large firms (SMB_t); the small and large stock portfolios include all stocks with market capitalization in the lower and upper deciles of the sample median;
- The difference between the expected returns on portfolios of stocks with high and low book-to-market ratios (HML_t).

Thus, the expected excess return of stock i can be represented as

$$R_{it} - r_{ft} = \beta_0 + \beta_1(R_{mt} - r_{ft}) + \beta_2SMB_t + \beta_3HML_t + u_t \quad (2.2.5)$$

As an example we consider monthly returns on IBM stocks for the period January 1990 to September 2007. This data with Fama-French factors is available as IBM1.xls. Variables in the data sets are

- *ibm* – monthly returns on IBM stocks;
- *Mkt* – monthly returns on the market index;
- *rf* – monthly rate of the risk-free rate;
- *SMB* and *HML* – Fama-French size and book-to-market risk factors, respectively.

In order to estimate the relation (2.2.5) we have to construct excess returns on IBM stocks and market portfolio. In EViews they can be created using

```
series ibm_ex=ibm-rf
series Mkt_ex=Mkt-rf
```

There are two ways of estimating linear regression in EViews. The first one, and more powerful, is through the main menu **Quick/Estimate Equation**. In the **Equation specification** window type the equation to be estimated. Using arithmetic operation we can specify the equation as

```
ibm_ex=C(1)+C(2)*Mkt_ex+C(3)*SMB+C(4)*HML
```

Note that coefficients of the equation should always be in the form C(1), C(2), etc. However, if the model is linear, it is more common to omit operation and coefficients signs and write

```
ibm_ex C Mkt_ex SMB HML
```

Note, that in the latter specification the dependent variable should be on the first place. The term **C** indicates that we are estimating the model with intercept; if it is omitted, the regression will be estimated without the intercept term.

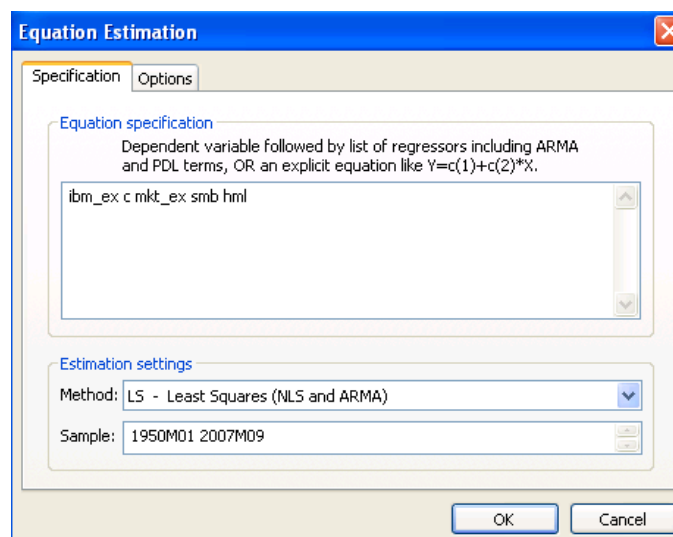


Figure 2.1: Regression estimation dialog window

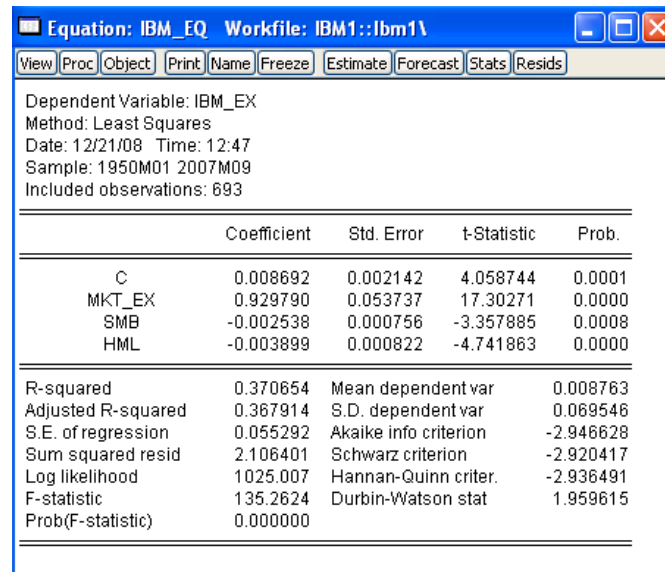
In the **Estimated setting window** make sure **LS – Least Squares (NLS and ARMA)** is chosen. The **Sample** window allows to estimate the model for different subsamples. This option subsample is specified in the same way as in the **Sample** object. Press **OK** and the regression output appears on the screen.

Another way of estimating a linear regression model is through the command line. To create an equation object use the declaration command **equation** following by a name of the object and the estimation type command (**ls** in our case stands for least squares) separated by the dot. Finally one should specify the model in the same way as above

```
equation ibmeq.ls ibm_ex C Mkt_ex SMB HML
```

Estimation Output The regression estimation output looks as follows

The estimated coefficients of the model are given in the column **Coefficients** (the coefficient in front of **C** denote estimate of the intercept term). Slope coefficients denote the sensitivities of the returns on the stock to the three factors and show the impact of systematic factors on returns. In column **t-statistic**, the value of the test statistic is provided to test that the hypothesis $\beta_i = 0$. All the coefficients are highly statistically significant as indicated by low p-values (column **Prob**). The



	Coefficient	Std. Error	t-Statistic	Prob.
C	0.008692	0.002142	4.058744	0.0001
MKT_EX	0.929790	0.053737	17.30271	0.0000
SMB	-0.002538	0.000756	-3.357885	0.0008
HML	-0.003899	0.000822	-4.741863	0.0000

R-squared	0.370654	Mean dependent var	0.008763
Adjusted R-squared	0.367914	S.D. dependent var	0.069546
S.E. of regression	0.055292	Akaike info criterion	-2.946628
Sum squared resid	2.106401	Schwarz criterion	-2.920417
Log likelihood	1025.007	Hannan-Quinn criter.	-2.936491
F-statistic	135.2624	Durbin-Watson stat	1.959615
Prob(F-statistic)	0.000000		

Figure 2.2: Regression estimation output

overall significance of the regression is reflected in the value of **F-statistic** which is high enough to reject the null hypothesis of insignificance of all slope coefficients (p-value is given in **Prob (F-statistic)**).



"I studied English for 16 years but...
...I finally learned to speak it in just six lessons"
Jane, Chinese architect

ENGLISH OUT THERE

Click to hear me talking before and after my unique course download

The proportion of the variance R_{it} explained by the variability in the market index is the usual regression R^2 statistic and $1 - R^2$ is the proportion of the variability of R_{it} that is due to firm specific factors. The proportion of market specific risk is $R^2 = 0.37$ and the proportion of firm specific risk is $1 - R^2 = 0.63$.

By estimating the regression model, EViews produces an object **Equation**, which can be saved and used later on (press **Name** button in the top of the equation window). As each object in EViews, **Equation** can be represented in different views. **View/Representation** view contains the equation specification of the model, **View/Estimation Output** provides the familiar model output. **View/Actual, Fitted, Residuals** creates various plots of the estimated residual series, as well as fitted values of the dependent variable. Residual series is automatically stored in the series object **resid** which created by EViews in each workfile. Note, that **resid** contains residuals of the last estimated model and will be lost once the model is reestimated. Thus, residual series has to be saved for further use, if necessary. This can done by copying the residual series into a new object

```
series resid_ibm=resid
```

Now, the residuals from the CAPM regression for IBM stock returns are stored in the new series object **resid_ibm**.

Besides the standard errors of the coefficient estimators, given in the output window, one can retrieve the whole variance-covariance matrix by clicking on **View/Covariance Matrix**.

Residuals Diagnostic Before drawing any conclusions from the estimated regression, it is necessary to perform residual diagnostic to make sure that the assumptions of the classic linear regression model are satisfied. This can be done in the section **View/Residual Tests. Correlogram - Q-statistic** provides values of the Box-Ljung statistics to test the significance of autocorrelations of residuals. The correlogram of the residuals from the factor model is given in Figure ??.

Low p-values indicate absence of serial autocorrelations up to lag 10. Another way to test for series correlation is to perform Breusch-Godfrey Test – in EViews this can be done through **Serial Correlation LM Test**. In the upper panel of the Breusch-Godfrey test output there are two versions of the test statistic which are asymptotically equivalent. Their p-values both confirm the absence of series autocorrelation up to the second order. The no-autocorrelation null hypothesis is also not rejected by the Durbin-Watson test; test statistic is given in the regression output is equal to 1.959 which is in the acceptance region.

The option **Histogram - Normality Test** builds the histogram of the residuals, their descriptive statistics as well as the value of the Jarque-Bera statistic.

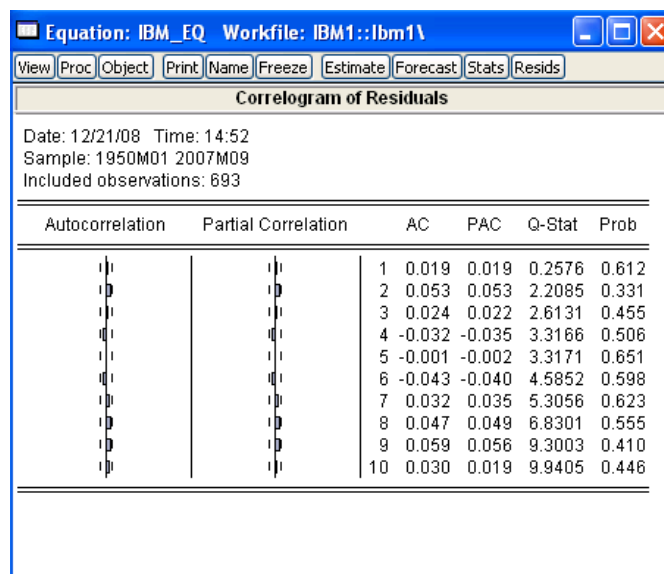


Figure 2.3: Correlogram of residuals from the factor model for IBM stock returns

Notice that the Jarque-Bera statistic indicates that the residuals from the CAPM regression are not normally distributed. Note that even the residuals are not normally distributed, the inference is still correct asymptotically.

EViews also provides a number of test to test the hypothesis of homoscedasticity on the regression. Under the option **Heteroscedasticity Tests...** one can choose among Breush-Godfrey-Pagan, Harvey, Glejser, ARCH and White tests.

Three of them – Breush-Godfrey-Pagan, Glejser and White tests reject the hypothesis of homoscedasticity while Harvey and ARCH test do not reject the null. The reason for using several tests is that there are many different possible alternatives for the form of heteroscedasticity.

All the tests for autocorrelations and heteroscedasticity can be performed through the command line as well.

For the Breusch-Godfrey test for serial correlation we should specify the name of the regression equation we need to test and then the command `auto(lags)` where *lags* corresponds to the order of autocorrelation being tested. For example

```
ibm_eq.auto(2)
```

will perform the test for second order autocorrelation in the factor model for IBM stock.

To perform heteroscedasticity tests we should specify the equation name followed by the command `hettest(options)`. In the *options* field we can specify the test being performed in the following way: `type=keyword`, where *keyword* is either "BPG"

(Breusch-Pagan-Godfrey - default), "Harvey", "Glejser", "ARCH", or "White". Inclusion of the command `c` in the options will lead to inclusion of cross-product terms in the auxiliary regression specification. Optionally, a list of variables may follow the command to include them into auxiliary regression as well.

For example,

```
ibm_eq.hettest(type="White",c)
```

will perform the White's test for heteroscedasticity for the `ibm_eq` equation.

Since the exact form of heteroscedasticity is not known, it is not clear how to perform GLS estimator in this case. EViews allows to compute heteroscedasticity consistent as well as heteroscedasticity and autocorrelation consistent coefficient covariance matrices. In order to compute them, click on **Estimate** button in the object menu and choose the **Options** tab. Tick the box in front of **Heteroscedasticity consistent coefficient matrix** to activate the option. Click **OK** to reestimate the model.

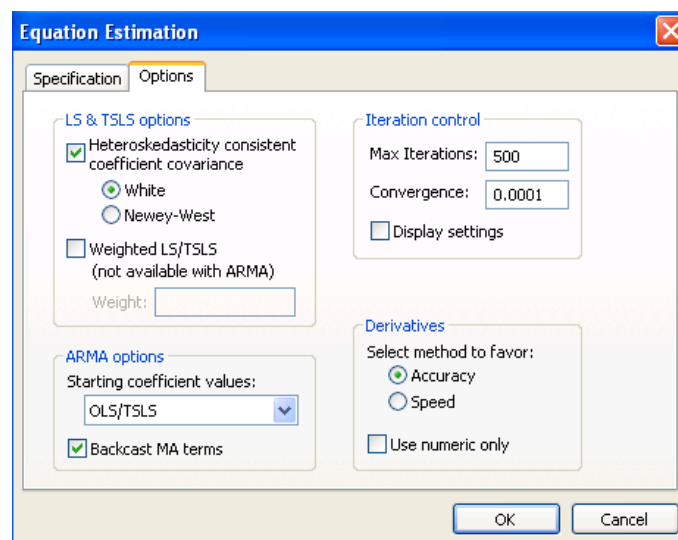


Figure 2.4: Regression output with White's heteroscedasticity adjusted standard errors

All coefficients remain still statistically significant using White's heteroscedasticity consistent standard errors.

Stability tests Finally, we can test the model for coefficients stability and structural breaks. In EViews this can be performed under the option **Views/Stability tests**. With Ramsey RESET test for model misspecification we cannot reject the null hypothesis of the correct specification (p-value 0.3754).

We start stability tests with the recursive residuals tests as they can help

us to detect visually potential breakpoints. Click on **Recursive Estimates (OLS only)** and choose **Recursive residuals**. EViews will produce the plot of recursively estimated residuals from the model together with their confidence intervals.

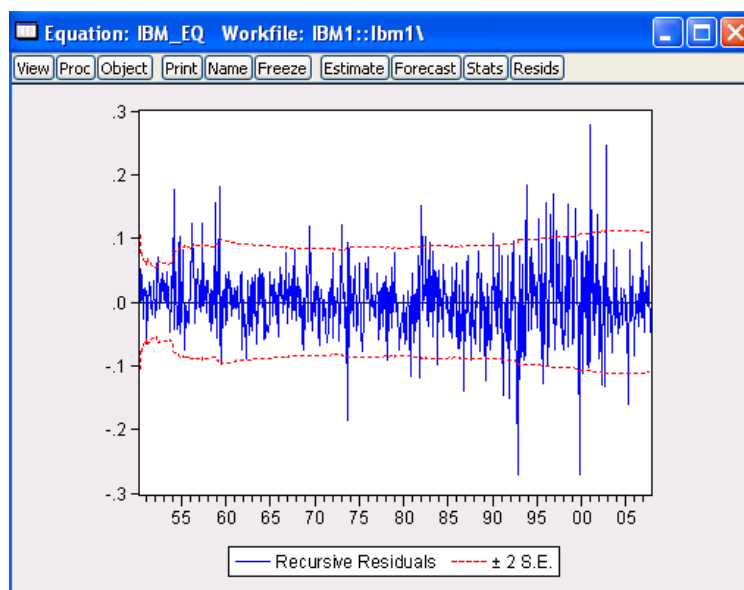


Figure 2.5: Recursively estimated residuals and their confidence bounds

What do you want to do?

No matter what you want out of your future career, an employer with a broad range of operations in a load of countries will always be the ticket. Working within the Volvo Group means more than 100,000 friends and colleagues in more than 185 countries all over the world. We offer graduates great career opportunities – check out the Career section at our web site www.volvogroup.com. We look forward to getting to know you!

VOLVO
AB Volvo (publ)
www.volvogroup.com

VOLVO TRUCKS | RENAULT TRUCKS | MACK TRUCKS | VOLVO BUSES | VOLVO CONSTRUCTION EQUIPMENT | VOLVO PENTA | VOLVO AERO | VOLVO IT
VOLVO FINANCIAL SERVICES | VOLVO 3P | VOLVO POWERTRAIN | VOLVO PARTS | VOLVO TECHNOLOGY | VOLVO LOGISTICS | BUSINESS AREA ASIA

Majority of the recursive residuals are within their confidence intervals however there are several outliers spraying out their bounds. These are potential points for the structural breaks in the models. The CUSUM test does not indicate any potential breakpoints, however the CUSUM squared test suggests that there may be some breaks in sixties and at the beginning of 2000.

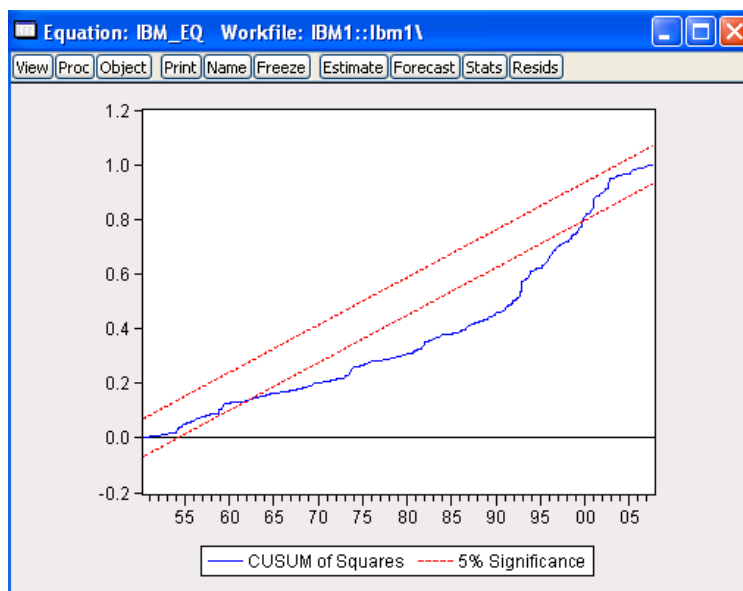


Figure 2.6: CUSUM squared statistics and its confidence bounds

We can go further and test whether there is a structural break at the specified dates using the Chow test. The p-value of the F-statistic for the Chow test is 0.3090 at the breakpoint January 1961 indicating no structural break at that date. However, if the breakpoint is specified at January 2000, we reject the null hypothesis of the parameters constancy at 1% significance level. Structural breaks may occur in the model due to some misspecifications. For example, from January 2000 there is one missing factor in the model which plays important role in explaining stock returns. The breakpoint data also corresponds to the dot.com bubble period where the classic factors model structure may change. In order to verify our hypothesis, we can include dummy variable corresponding to the bubble period to eliminate the effect from the model. To create the dummy which is equal to 1 for the period from January 2000 to December 2001, we write

```
series dummy=0
smpl 2000M01 2001M12
series dummy=1
```


smpl @all

Since structural breaks may occur in all the parameters, we include the dummy variable interacting with all regressors. Thus, in the **Estimate equation** box we have to specify a new model

ibm_ex C Mkt_ex SMB HML dummy*Mkt_ex dummy*SMB dummy*HML

As a result, the coefficient for interacting term with market portfolio returns and SMB factor is insignificant, however the interacting term with HML factor is statistically significant at 10% level.

The screenshot shows the EViews regression output window for the equation 'UNTITLED' in the workfile 'IBM1::ibm1'. The dependent variable is 'IBM_EX' and the method is 'Least Squares'. The sample period is from 2000M01 to 2007M09, with 93 observations included. The output table provides coefficients, standard errors, t-statistics, and probabilities for each parameter. It also includes summary statistics such as R-squared, adjusted R-squared, and various diagnostic test statistics.

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.013473	0.006911	1.949517	0.0545
MKT_EX	1.508672	0.231477	6.517579	0.0000
SMB	-0.007286	0.003156	-2.308490	0.0234
HML	-0.013625	0.003824	-3.562766	0.0006
DUMMY*MKT_EX	-0.434852	0.390015	-1.114963	0.2680
DUMMY*SMB	0.003600	0.004081	0.882275	0.3801
DUMMY*HML	0.009399	0.005163	1.820551	0.0722

R-squared	0.555885	Mean dependent var	0.002823
Adjusted R-squared	0.524900	S.D. dependent var	0.089562
S.E. of regression	0.061733	Akaike info criterion	-2.659709
Sum squared resid	0.327743	Schwarz criterion	-2.469083
Log likelihood	130.6765	Hannan-Quinn criter.	-2.582740
F-statistic	17.94059	Durbin-Watson stat	2.264644
Prob(F-statistic)	0.000000		

Figure 2.7: Output of the regression estimation with dummy variables

Correcting of misspecification also helps to improve properties of residuals. After introducing dummy variable all tests for heteroscedasticity indicate either no heteroscedasticity or produce some marginally significant p-values.

Testing linear restrictions EViews makes it possible to test hypothesis on the coefficient restrictions by means of Wald test. Consider testing the joint null hypothesis $\beta_1 = 1$ and $\beta_2 = \beta_3$. This hypothesis imposes two linear restrictions on the parameter vector. In the **View** option of the object menu choose **Coefficient Tests/Wald – Coefficient Restrictions....** Type the restrictions to be tested in the box. Note that coefficients of the model are denoted by $C(1)$, $C(2)$, etc. In order to find out the exact notations of the parameters, go to **View/Representation**. P-value of the Wald test statistic is higher than any reasonable significance levels so we do not reject the null hypothesis of the validity of the restrictions.

The Wald test can also be performed in the command line. One should first specify the name of equation being tested followed by dot and command `wald`. Specifications of the restrictions follows separated by commas.

```
eq1.wald c(2)=0, c(3)=0
```

Predictions After having estimated the regression, often our aim is to construct forecast of the dependent variable. EViews' forecast function can be invoked through **Forecast** option in the menu of the equation object. In the box **Forecast name** type the name of the variable where the regression forecasts will be stored. EViews will automatically create a new series object with the specified name and plot the predicted series with two confidence region bounds.

Alternatively, one can view the forecast by double-click on the forecast variable. In the menu option **View** choose **Graph** where the required graph type can be generated.

In order to generate forecasts through the command line, use the command `fit` followed by a name of a series variable where the forecast values should be stored

```
ibmeq.fit ibm_exf
```

qaiTEYE[®]
Challenge the way we run

**EXPERIENCE THE POWER OF
FULL ENGAGEMENT...**

**RUN FASTER.
RUN LONGER..
RUN EASIER...**

**READ MORE & PRE-ORDER TODAY
WWW.GAITEYE.COM**

EViews also allows to generate standard errors of the forecasts along with the predictions themselves. To do this, simply include a name of another variable at the end of the line.

```
ibmeq.fit ibm_exf ibm_exfst
```

2.2.4 Programming Example

Note that the factor sensitivities of stock (portfolio) returns represented by the estimated coefficients vary through time. As the model is estimated for alternative sample periods, the estimated coefficients will change. A useful analogy is the value of a stock's beta that varies through time based on the sample period data used to estimate the security market line.

The estimated factor model (2.2.5) for IBM uses all of the data over the 57 year period from January 1950 to September 2007. It is generally thought that coefficients do not stay constant over such a long time period. To take into account this fact while building returns forecast based on the factor model, we can perform rolling window regression. We start with initializing necessary variables (e.g., number of observation in the workfile, length of the window). For this purpose we make sure that the current sample is set to the whole range of the data. Type the following commands in a new program window:

```
smpl @all
scalar n=@obs(ibm_ex)
scalar window=60
```

Next, we create new object we will be using in the program – series of the forecasts and an equation object.

```
series ibm_exf
equation e
```

In the next lines we specify a loop where we reset the current sample to the estimation window and roll it across the data range. For each of the subsamples we estimate the factor model.

```
for !i=0 to n-window-1
smpl @first+!i @first+!i+window-1
e.ls ibm_ex c Mkt_ex SMB HML
```

Once the model is estimated we reset the sample to a subsample where we want

to forecast returns. Since we build one-step-ahead forecast, our new subsample will be just one observation ahead the estimation window.

```
smpl @first+!i+window @first+!i+window
```

We generate the forecast using the estimated model. To access the values of the estimated parameters we use EViews function `@coefs`. In parentheses we specify the order of the parameter – it corresponds to the order of respective variable in the regression model. Note that `@coefs` contains the values of the last estimated model. Once the equation is re-estimated, the new values of parameters are stored in `@coefs`.

```
ibm_exf=@coefs(1)+@coefs(2)*Mkt_ex+@coefs(3)*smb+@coefs(4)*hml
```

```
next
```

Just to tidy up the workfile we delete auxiliary variables `window` and `n`.

```
delete window n
```

The series `ibm_exf` contains the generated forecast from the rolling window model.

Similarly to the use of `@coefs` function, one can access other OLS statistics. The specifications are given in Table 2.2.

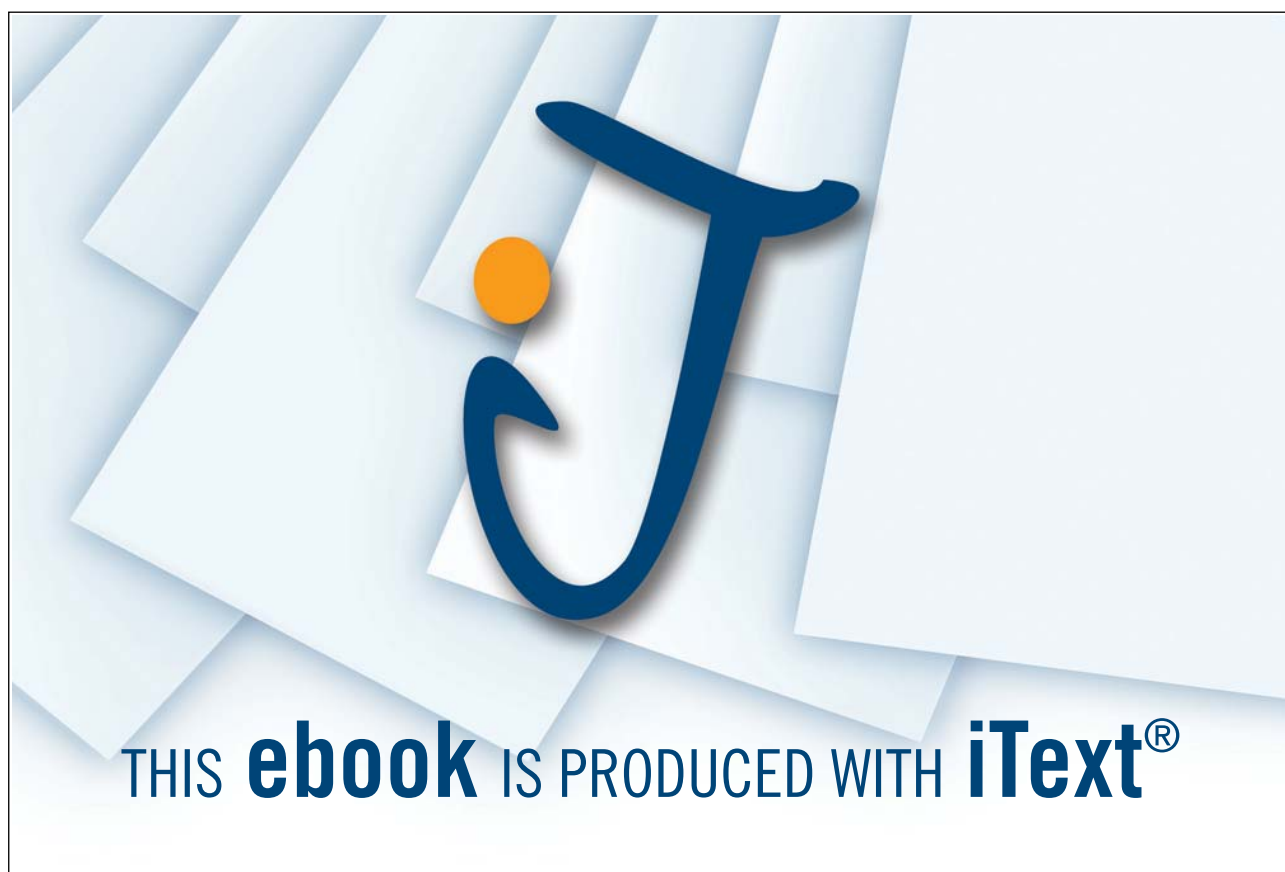


Table 2.2: Equation Data Members

Data Member	Description
@aic	Akaike information criterion
@coefcov(i,j)	covariance of coefficient estimates β_i and β_j
@coefs(i)	i-th coefficient value
@dw	Durbin-Watson statistic
@f	F-statistic
@hq	Hannan-Quinn information criterion
@logl	value of the log likelihood function
@meandep	mean of the dependent variable
@ncoef	number of estimated coefficients
@r2	R-squared statistic
@rbar2	adjusted R-squared statistic
@regobs	number of observations in regression
@schwarz	Schwarz information criterion
@sddep	standard deviation of the dependent variable
@se	standard error of the regression
@ssr	sum of squared residuals
@stderrs(i)	standard error for coefficient
@tstats(i)	t-statistic value for coefficient
@coefcov	matrix containing the coefficient covariance matrix
@coefs	vector of coefficient values
@stderrs	vector of standard errors for the coefficients
@tstats	vector of t-statistic values for coefficients

2.3 Nonlinear Regression

In many cases the relation between variables can happen to be nonlinear. If such model cannot be transformed into a linear one, we call such model intrinsically nonlinear regression model.

We can represent such model in the following way

$$bfY = F(\mathbf{X}, \theta) + \mathbf{u},$$

where F is a non-linear function, where $\mathbf{X}_i = [1, X_{2i}, \dots, X_{ki}]'$ is a $k \times 1$ vector of explanatory variables, and u_i is a random error term.

The least squares estimation problem to minimize

$$S(\theta) = \sum_{i=1}^n (Y_i - F(X_i, \theta))^2 \quad (2.3.1)$$

becomes non-linear. The first order conditions are given by


$$\frac{\partial S(\theta)}{\partial \theta_j} = -2 \sum_{i=1}^n (Y_i - F(X_i, \theta)) \frac{\partial F}{\partial \theta_j}.$$

This gives a set of non-linear normal equations in θ . The non-linear least squares (NLS) estimator $\hat{\theta}_{NLS}$ is defined as the minimizing value of (2.3.1).

In EViews, the Nonlinear Least Squares method has the same implementation as the OLS. The only difference is that the model in the **Equation specification** box should be entered as a mathematical expression instead of a list of variables. For example,

$$y = \exp(c(1)*x) + (c(2)*z + 4)^2$$

Interpretation of the estimation output, residual diagnostic and inference can be performed in the same way as for the OLS regression.



www.sylvania.com

We do not reinvent
the wheel we reinvent
light.

Fascinating lighting offers an infinite spectrum of possibilities: Innovative technologies and new markets provide both opportunities and challenges. An environment in which your expertise is in high demand. Enjoy the supportive working atmosphere within our global group and benefit from international career paths. Implement sustainable ideas in close cooperation with other specialists and contribute to influencing our future. Come and join us in reinventing light every day.

Light is OSRAM

OSRAM
SYLVANIA