

# Advanced Forecasting Techniques and Models: Regression & Econometrics

Short Examples Series  
using  
Risk Simulator



For more information please visit:  
[www.realoptionsvaluation.com](http://www.realoptionsvaluation.com)  
or contact us at:  
[admin@realoptionsvaluation.com](mailto:admin@realoptionsvaluation.com)

## ***Forecasting – Econometric, Correlations and Multiple Regression Modeling***

**File Name:** *Forecasting – Econometric, Correlations and Multiple Regression Modeling*

**Location:** *Modeling Toolkit | Forecasting | Econometric, Correlations and Multiple Regression Modeling*

**Brief Description:** *This sample model illustrates how to use Risk Simulator for running a multiple regression analysis and econometric modeling*

**Requirements:** *Modeling Toolkit, Risk Simulator*

This example shows how a multiple regression can be run using Risk Simulator. The raw data are arranged in the *Cross-Sectional Data* worksheet, which contains cross-sectional data on 50 months of a bank's *key risk indicators* as they pertain to Basel II's Operational Risk analysis guidance. The idea is to see if there is a relationship between the *Impact Losses (\$M)* per month and these explanatory variables using multiple regression, econometric modeling and correlation analysis.

### **Procedure**

To run this model, simply:

1. In the *Cross Sectional Data* worksheet, select the area **C5:H55** (Figure 1).
2. **Select Risk Simulator | Forecasting | Multiple Regression.**
3. Choose *Impact Losses* as the dependent variable in the regression and click on **OK** to run the regression (Figure 2).

### **Model Results Analysis**

Note that more advanced regressions such as lag regressors, stepwise regression, and nonlinear regressions can also be run using Risk Simulator. For more details on running such regressions as well as the interpretation of the regression report, refer to *Modeling Risk, 2nd Edition* (Wiley 2006) by Dr. Johnathan Mun.

### **Results Summary**

Refer to the *Report* worksheet for details on the regression output. The report has more details on the interpretation of specific statistical results. It provides these elements: multiple regression and analysis of variance output, including coefficients of determination, hypothesis test results (single variable t-test and multiple variable F-test), computed coefficients for each regressor, fitted chart, and much more.

Additional econometric analyses can be performed. For instance, a nonlinear relationship is appropriate here. Simply follow the three-step procedure described, but remember to check the **Nonlinear Regression** box before hitting **OK**. Also, interactions between variables might exist in this model, and a

modification of the data variables can be done before further regression models are run. In such instances, combine various independent variables to create new variables (e.g.,  $X_1$ ,  $X_2$ , ...  $X_n$ , and  $X_1X_2$ ,  $X_1X_3$ ,  $X_1X_4$ , and so forth, and then run a **Stepwise Regression**).

Finally, the chapter on *Forecasting – Data Diagnostics* provides the critical modeling prerequisites before running any multivariate regression analysis, to make sure that all the required assumptions have not been violated in the data set, and to check the modeling and predictive quality of the data.

### Multiple Regression and Econometric Analysis Data Set

Impact Losses (\$M)	Cycle Time and Timeliness of Transactions	Transaction Volume	Hiring and Training Costs	Customer Satisfaction Index	IT Network Downtime
521	18308	185	4.041	79.6	7.2
367	1148	600	0.55	1	8.5
443	18068	372	3.665	32.3	5.7
365	7729	142	2.351	45.1	7.3
614	100484	432	29.76	190.8	7.5
385	16728	290	3.294	31.8	5
286	14630	346	3.287	678.4	6.7
397	4008	328	0.666	340.8	6.2
764	38927	354	12.938	239.6	7.3
427	22322	266	6.478	111.9	5
153	3711	320	1.108	172.5	2.8
231	3136	197	1.007	12.2	6.1
524	50508	266	11.431	205.6	7.1
328	28886	173	5.544	154.6	5.9
240	16996	190	2.777	49.7	4.6
286	13035	239	2.478	30.3	4.4
285	12973	190	3.685	92.8	7.4
569	16309	241	4.22	96.9	7.1
96	5227	189	1.228	39.8	7.5
498	19235	358	4.781	489.2	5.9
481	44487	315	6.016	767.6	9
468	44213	303	9.295	163.6	9.2
177	23619	228	4.375	55	5.1
198	9106	134	2.573	54.9	8.6
458	24917	189	5.117	74.3	6.6
108	3872	196	0.799	5.5	6.9
246	8945	183	1.578	20.5	2.7
291	2373	417	1.202	10.9	5.5
68	7128	233	1.109	123.7	7.2
311	23624	349	7.73	1042	6.6
606	5242	284	1.515	12.5	6.9
512	92629	499	17.99	381	7.2

Figure 1: Econometric data set

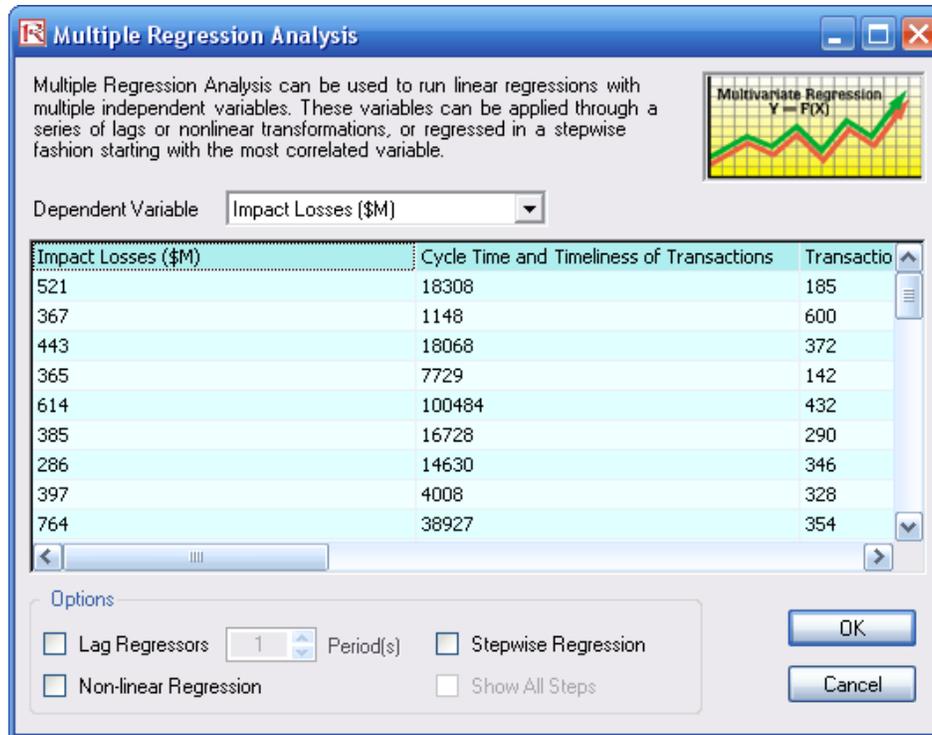


Figure 2: Running a multiple regression

Sometimes, the variables themselves require tweaking before a multiple regression is run. This is the basics of econometrics. Using Risk Simulator, this process can be applied quickly and the variables can be manipulated (log, addition, subtraction, multiplication, lagged, and so forth) and modeled. The procedure is described next.

### Procedure

To run a basic econometrics model, simply:

1. In the *Cross Sectional Data* worksheet, select the area **C5:H55** (Figure 1).
2. **Select Risk Simulator | Forecasting | Basic Econometrics.**
3. Type in the required single dependent variable and multiple independent variables with the required modifications but separated by a semi-colon (Figure 3). Click **Show Results** to view the results of this model, and make additional modifications as required, and when done, click on **OK** to generate the full model report as seen in Figure 4.

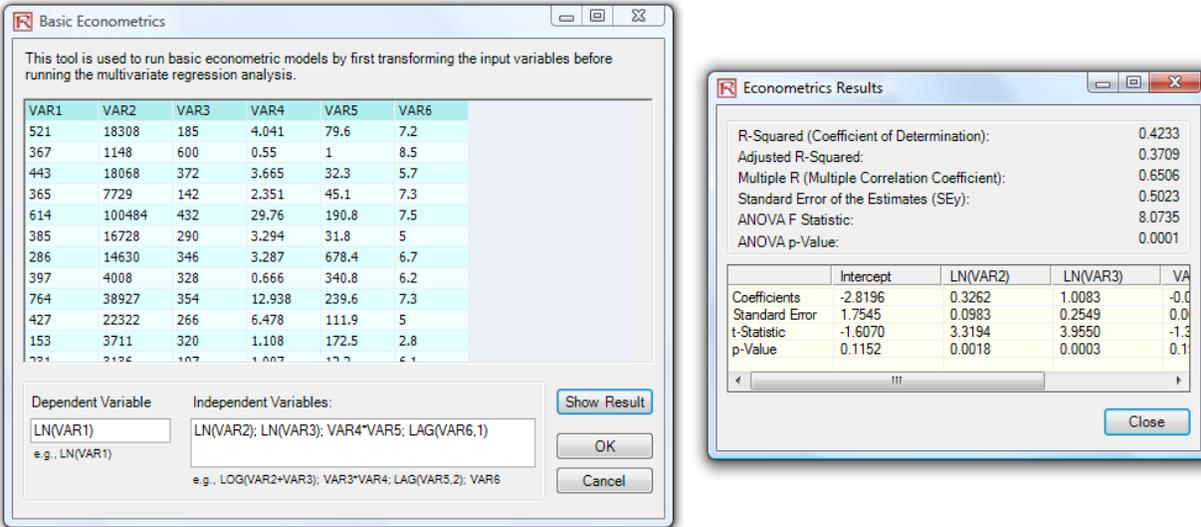


Figure 3: Basic econometrics results

### Basic Econometrics Analysis Report

#### Regression Statistics

R-Squared (Coefficient of Determination)	0.4233
Adjusted R-Squared	0.3709
Multiple R (Multiple Correlation Coefficient)	0.6506
Standard Error of the Estimates (SEy)	0.5023
Number of Observations	49

The R-Squared or Coefficient of Determination indicates that 0.42 of the variation in the dependent variable can be explained and accounted for by the independent variables in this regression analysis. However, in a multiple regression, the Adjusted R-Squared takes into account the existence of additional independent variables or regressors and adjusts this R-Squared value to a more accurate view of the regression's explanatory power. Hence, only 0.37 of the variation in the dependent variable can be explained by the regressors.

The Multiple Correlation Coefficient (Multiple R) measures the correlation between the actual dependent variable (Y) and the estimated or fitted (Y) based on the regression equation. This is also the square root of the Coefficient of Determination (R-Squared).

The Standard Error of the Estimates (SEy) describes the dispersion of data points above and below the regression line or plane. This value is used as part of the calculation to obtain the confidence interval of the estimates later.

#### Regression Results

	Intercept	LN(VAR2)	LN(VAR3)	VAR4*VAR5	LAG(VAR6,1)
Coefficients	-2.8196	0.3262	1.0083	-0.0001	-0.0209
Standard Error	1.7545	0.0983	0.2549	0.0001	0.0516
t-Statistic	-1.6070	3.3194	3.9550	-1.3271	-0.4055
p-Value	0.1152	<b>0.0018</b>	<b>0.0003</b>	<b>0.1913</b>	<b>0.6871</b>
Lower 5%	-6.3556	0.1282	0.4945	-0.0002	-0.1248
Upper 95%	0.7165	0.5243	1.5221	0.0000	0.0830

#### Degrees of Freedom

Degrees of Freedom for Regression	4
Degrees of Freedom for Residual	44
Total Degrees of Freedom	48

#### Hypothesis Test

Critical t-Statistic (99% confidence with df of 44)	2.6923
Critical t-Statistic (95% confidence with df of 44)	2.0154
Critical t-Statistic (90% confidence with df of 44)	1.6802

The Coefficients provide the estimated regression intercept and slopes. For instance, the coefficients are estimates of the true; population b values in the following regression equation  $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_nX_n$ . The Standard Error measures how accurate the predicted Coefficients are, and the t-Statistics are the ratios of each predicted Coefficient to its Standard Error.

The t-Statistic is used in hypothesis testing, where we set the null hypothesis (Ho) such that the real mean of the Coefficient = 0, and the alternate hypothesis (Ha) such that the real mean of the Coefficient is not equal to 0. A t-test is performed and the calculated t-Statistic is compared to the critical values at the relevant Degrees of Freedom for Residual. The t-test is very important as it calculates if each of the coefficients is statistically significant in the presence of the other regressors. This means that the t-test statistically verifies whether a regressor or independent variable should remain in the regression or it should be dropped.

The Coefficient is statistically significant if its calculated t-Statistic exceeds the Critical t-Statistic at the relevant degrees of freedom (df). The three main confidence levels used to test for significance are 90%, 95% and 99%. If a Coefficient's t-Statistic exceeds the Critical level, it is considered statistically significant. Alternatively, the p-Value calculates each t-Statistic's probability of occurrence, which means that the smaller the p-Value, the more significant the Coefficient. The usual significant levels for the p-Value are 0.01, 0.05, and 0.10, corresponding to the 99%, 95%, and 90% confidence levels.

The Coefficients with their p-Values highlighted in blue indicate that they are statistically significant at the 90% confidence or 0.10 alpha level, while those highlighted in red indicate that they are not statistically significant at any other alpha levels.

## Analysis of Variance

	Sums of Squares	Mean of Squares	F-Statistic	p-Value	Hypothesis Test	
Regression	8.15	2.04	8.07	0.0001	Critical F-statistic (99% confidence with df of 4 and 44)	3.7784
Residual	11.10	0.25			Critical F-statistic (95% confidence with df of 4 and 44)	2.5837
Total	19.25				Critical F-statistic (90% confidence with df of 4 and 44)	2.0772

The Analysis of Variance (ANOVA) table provides an F-test of the regression model's overall statistical significance. Instead of looking at individual regressors as in the t-test, the F-test looks at all the estimated Coefficients' statistical properties. The F-Statistic is calculated as the ratio of the Regression's Mean of Squares to the Residual's Mean of Squares. The numerator measures how much of the regression is explained, while the denominator measures how much is unexplained. Hence, the larger the F-Statistic, the more significant the model. The corresponding p-Value is calculated to test the null hypothesis (Ho) where all the Coefficients are simultaneously equal to zero, versus the alternate hypothesis (Ha) that they are all simultaneously different from zero, indicating a significant overall regression model. If the p-Value is smaller than the 0.01, 0.05, or 0.10 alpha significance, then the regression is significant. The same approach can be applied to the F-Statistic by comparing the calculated F-Statistic with the critical F values at various significance levels.

## Forecasting

Period	Actual (Y)	Forecast (F)	Error (E)
1	5.9053618	5.7783	0.1270
2	6.0935698	6.1590	(0.0654)
3	5.8998974	4.9704	0.9295
4	6.4199949	6.4560	(0.0360)
5	5.9532433	5.9047	0.0485
6	5.6559918	5.9237	(0.2677)
7	5.9839363	5.5699	0.4141
8	6.6385678	6.1722	0.4663
9	6.056784	5.8670	0.1898
10	5.0304379	5.5582	(0.5278)
11	5.4424177	5.0743	0.3682
12	6.2614917	6.0303	0.2312
13	5.7930136	5.5111	0.2819
14	5.4806389	5.5144	(0.0337)
15	5.6559918	5.6913	(0.0353)
16	5.6524892	5.4415	0.2110
17	6.3438804	5.6879	0.6559
18	4.5643482	5.1063	(0.5420)
19	6.2106001	5.9865	0.2241
20	6.1758673	5.9846	0.1912
21	6.1484683	6.1230	0.0255
22	5.1761497	5.7285	(0.5523)
23	5.288267	4.9752	0.3131
24	6.1268692	5.5583	0.5686
25	4.6821312	5.0591	(0.3769)
26	5.5053315	5.2546	0.2508
27	5.6733233	5.7414	(0.0681)
28	4.2195077	5.4450	(1.2255)

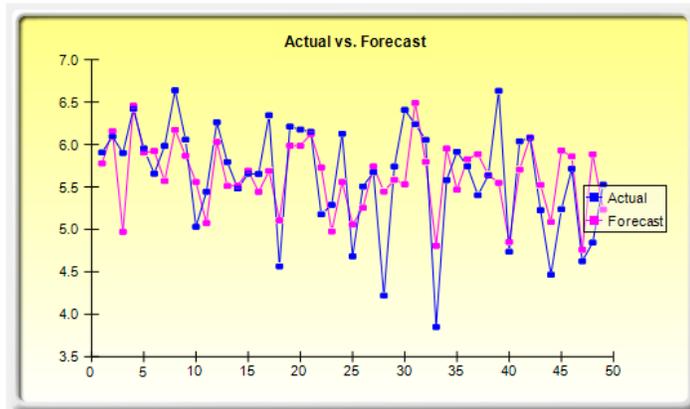


Figure 4: Basic econometrics report