

Chapter 5

Parameter Estimation:

MODLER's regression commands at their most basic are essentially intuitive. For example, consider:

$$\text{IMP}=\text{F}(\text{GNP},\text{CAPI})$$

which specifies that IMP is a function F() of the variables GNP and CAPI. This is the standard form: to execute a linear-in-parameters regression, simply create an expression that states that the dependent variable (left-hand-side) is a function of one or more independent (right-hand-side) variables. If you issue this command, provided that you have opened the demonstration data bank and set the date range and observation frequency, MODLER will find the relevant series, perform the regression, and display the results shown in Figure 5-1.

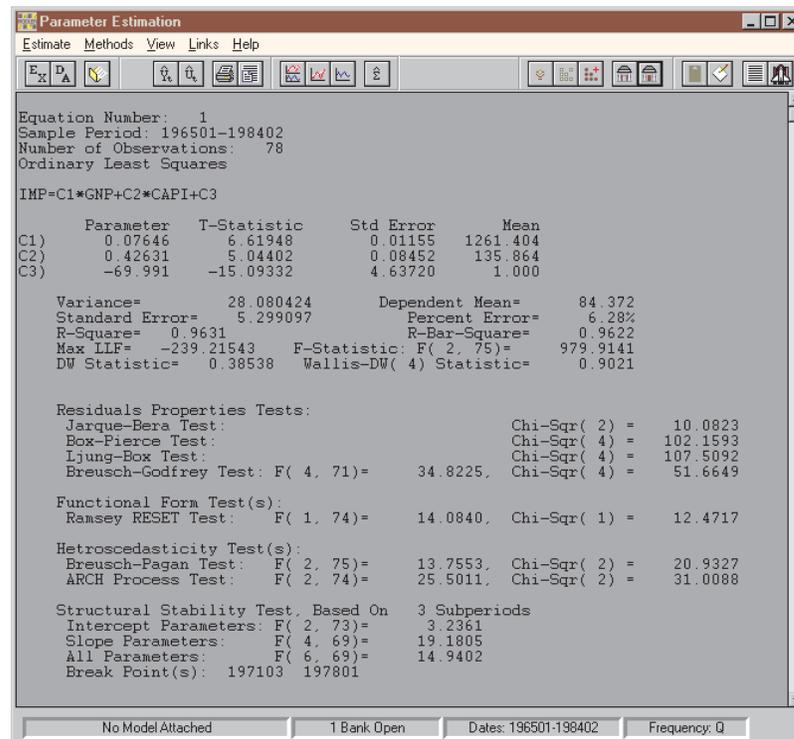


Figure 5-1. Regression Results Example

More generally, the extended form of the Ordinary Least Squares regression command is:

$$Y=F(\text{term}_1,\text{term}_2,\dots,\text{term}_k;\text{qualifiers})$$

where the terms themselves can be variables or expressions and the qualifiers—separated from the regressor terms by a semi-colon, as shown—further define the characteristics of the regression, as will be described.

However, before considering this command and its results in detail, it may be useful to know the context in which it can be used. There are 3 ways to issue a regression command. The first is as an ordinary command in the context of the command space of the Central Control screen. Alternatively, if you click on the Central Control screen menu category **Analysis**, and choose **Regression**, you will invoke the Parameter Estimation Screen and an initial display something like that shown in Figure 5-2. The third method, when building a multi-equation model, will be addressed in the next chapter.

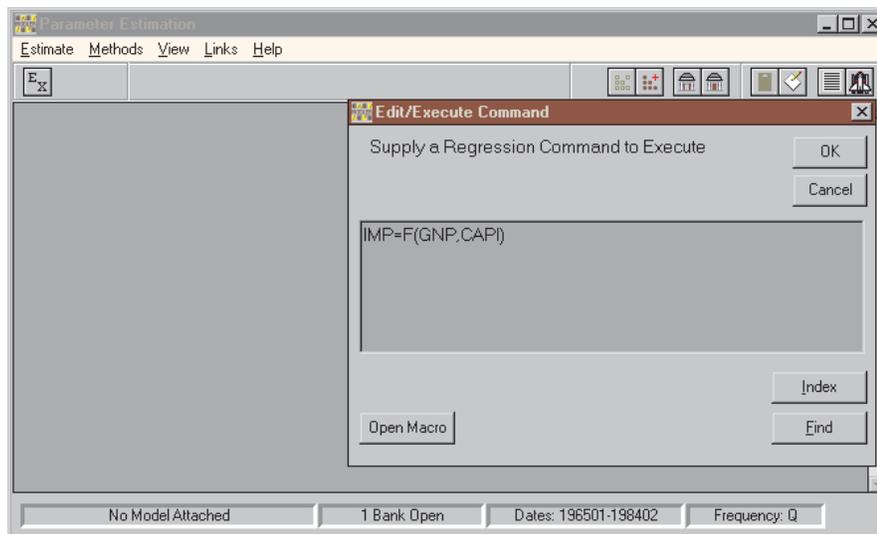
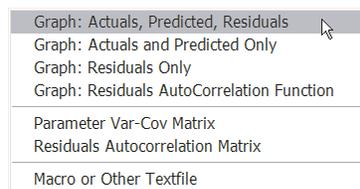


Figure 5-2. Parameter Estimation Screen Regression

Considering Figure 5-2, note, in particular, the **Index**, **Find**, and **Open Macro** buttons. Collectively, these provide information about available data series and allow you to capture mnemonics and possibly even entire regression commands from an existing macro file, for use in forming a regression command.

Supplementary Regression Commands

There are also several supplementary commands and menu item equivalents that provide further information about a just-executed regression. For example, you can display the parameter estimator variance-covariance matrix by clicking on the menu category **View**, which offers the choice list:



Choosing the last of these, when there are two regressors and a constant term, the parameter variance-covariance matrix takes the form:

PARAMETER VARIANCE-COVARIANCE MATRIX:

	1	2	3
1	.000133		
2	-.000942	.007143	
3	-.040256	.218121	21.503640

Predicted Values and Residuals

Alternatively, to display a comparison of actual and ‘predicted’ values of the dependent variable, a possible choice is the first **Graph: Actuals, Predicted, Residuals**, which produces the plot shown in Figure 5-3, displaying both the actual and predicted values of the dependent variable and the residuals.

This regression plot has a number of the characteristics of an ordinary plot. For instance, you are able to edit it using the dropdown menu system shown in the figure and described in Chapter 2 or else capture it to the Clipboard, among other obvious options. In addition, you can generate a macro file that will replicate both the regression and graph; because Graph is a subsidiary regression option, this macro does not in fact directly reproduce the plot, but

110 Parameter Estimation

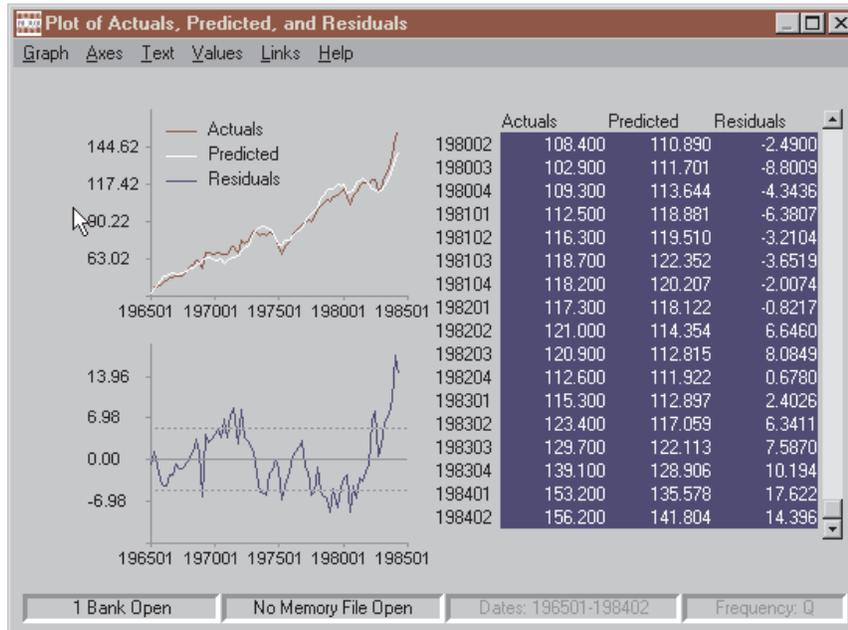


Figure 5-3. Sample Regression Plot

instead the regression, at which point the GRAPH option can again be chosen.

You can also create regression plots from scratch: to do this, first store the predicted and residual values in the Memory File either by clicking on the **Estimate** menu category of the Parameter Estimation Screen, which displays the choice list:

Edit/Execute
Drop-Add
AutoCode Equation (Puteq)
Attach Model
Create New Model ▶
Open Bank
Close Bank
Save Residuals
Save Sample Period Predicted
Save Last Regression Command
Edit Display Using Internal TextEditor
Print Display to Printer
Copy Display to WINWORD
Quit

from which you can make the appropriate choices. Or you can use the commands:

```
SAVE PREDICTED AS seriesname
SAVE RESIDUALS AS seriesname
```

in a macro or from the Central Control screen. So long as no other operation is pending, you can return from the Parameter Estimation screen to the Central Control screen simply by pressing the <Esc> key, after which you can issue these commands.

The ability to save these results makes it possible to treat the predicted and residuals just like any other time series. This capability allows you to build your own, specially tailored, plots of predicted or residual values, instead of or in addition to those produced via the **View** menu choices. For example, the commands:

```
SAVE PREDICTED AS PRED
PLOTTITLE:Graph of Actual and Predicted Regression Values
PLOT IMP,PRED
```

together will produce a plot of actuals and predicted.

Alternatively, consider:

```
SAVE RESIDUALS AS RESIDLS
HIGH=SDEV(RESIDLS)
LOW=-HIGH
PLOT RESIDLS,HIGH,LOW,0
```

Together these create a plot that shows the regression residuals and the points at which the values are within or greater than one standard deviation—bear in mind that the sum of the residuals is zero, at least so long as a constant term appears in the regression.

Incidentally, there is an important difference between these last two plots and those produced by clicking on the **View** menu category of the Parameter

112 Parameter Estimation

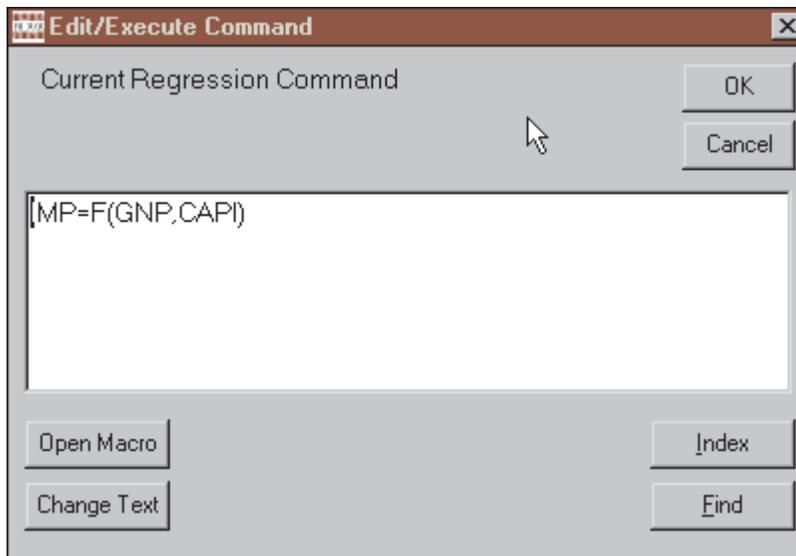
Estimation Screen: the plots last produced have all the properties of ordinary plots; that is, they are generated from data series retrieved from the Memory File or a data bank. They therefore can be produced from a set of explicit plot commands. In contrast, the regression Graph option plot is directly generated from the regression results as a special display.

Other Important Regression-related options

Look back at the choice list shown on page 104, which you will recall is displayed by clicking on the **Estimate** menu category. In addition to the commands just described that save residuals and predicted values, this list includes several others that manage the parameter estimation process. Among these are selections that permit the editing of regression commands, banks to be opened or closed, and results to be printed.

Edit/Execute Command

The **Edit/Execute** choice in particular recalls the command for the current regression and permits it to be directly edited in a variety of ways. Once a regression has been run there are two ways to invoke this option. The first is to click on **Edit/Execute**. Alternatively, while the regression results are displayed, you can point your cursor at these and left-click the mouse. You should next see a dialogue box that displays the regression command:



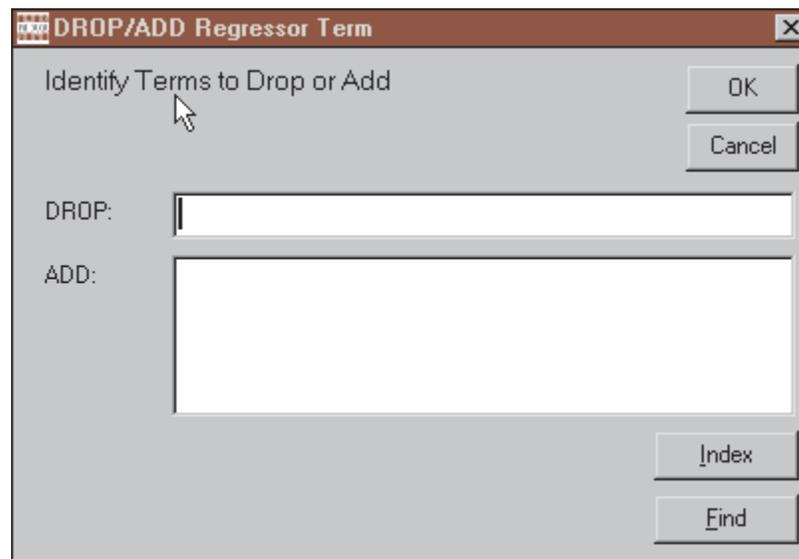
Note that the white space that displays this command is meant to contain text; one of the implications is that if you click on this space, using your *right* mouse button, and if you are using Windows 95 or a later operating system, a floating text edit menu will appear. Note in addition the **Change Text** button on the lower left of the form. You will also find that if you click your mouse on the text of the command, the backspace, delete, and other editing keys work. In short, you have the ability to edit the text in any way you see fit.

As a bonus, the **Open Macro** button can be used to gain access to regression commands and other text that you may find relevant. And once you open a macro, if you highlight a segment of text there, it will be copied immediately to the Edit/Execute command space.

In turn, the **Index** and **Find** buttons can be used to discover variables in open data banks or the Memoryfile that you may wish to use in regression commands.

Drop-Add Command

The Drop-Add facility is more specific in its function. Click on **Drop-Add** and you should immediately see the dialogue box:



This facility refers to regressor variables *only*; changes in the dependent variable must be made using the Edit/Execute command. Within this restric-

tion, the terms to be dropped from the existing regression are identified by number, counting from the left-most regressor term. Up to the number of regressor terms, you can include as many numbers as you wish in the text space following the word DROP. In the text space following the word ADD, you can include new regressor terms, separated by commas; these regressor terms can be variable mnemonics or expressions. The **Index** and **Find** buttons are provided in order to assist you when *Adding* additional terms.

Saving Estimated Equations: Single Equation Model

MODLER differs from an ordinary regression package in a number of ways, but a characteristic difference is that it allows you to save an estimated equation either as a functioning single equation model or as one of the equations in an existing multi-equation model file.

Considering first the Single Equation case, in order to create a model from an equation you have just estimated, simply click on **Create New Model** and then make the choice **Single Equation Model**. A form will appear, containing several text spaces to be filled in. Using this form, you will be asked to name the model, describe it if you wish, and specify the frequency of the data to be used with it. In addition, if relevant, you may need to specify the normalization, the implications of which are further discussed below.

Once you have identified the model and its characteristics, the Parameter Estimation Screen menu will immediately change. It will now include an additional category, **Projections**, as shown in Figure 5-4.

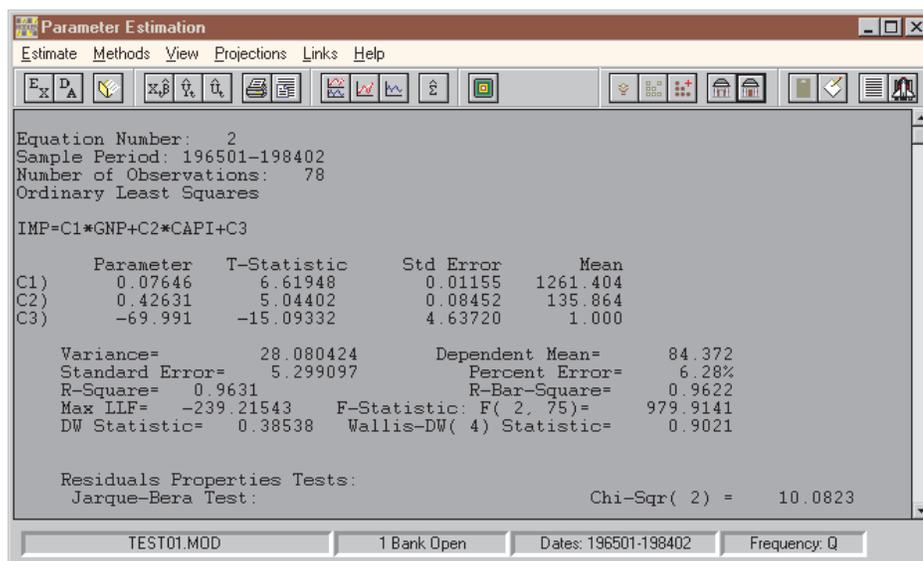


Figure 5-4. Estimation Screen with Projections

If you then click on **Projections**, you should see immediately a choice list in which only the first item will be enabled, **Project Equation**. The other items refer to the post-projection display of results, so are not relevant until an initial projection has been made. Clicking on **Project Equation** will cause another form to be displayed:

This form specifies the regression sample period essentially as a reminder, for reference as you choose the projection period. The Projections Results Details in turn identify the file used for the model and the datasource from which the assumptions data are to be retrieved; unless you change it, the datasource will be that used for the regression. The Projected Values Mnemonic is the name under which the projections results will be stored in the Memoryfile; note that each projection will presumably result in a new set of projected values, so that in general you should change this mnemonic from one projection to the next—if you use the same name 2 or more times, previous values stored under the given name will be overwritten without warning.

The **Project** button causes the projection to be made for the projection period specified. **Model** provides a description of the model's characteristics. **Assumptions** allows you to change the values of the regressor variables prior to a projection. Changes to regressor variable values are stored in the Memoryfile, not in the bank from which they are originally obtained; the effect is to preserve the original data. **Help** provides reasonably detailed context specific help, including how to display Single Equation model projections, which therefore will not be described here.

An aspect of this process that you should keep foremost in mind is that any variables stored in the Memoryfile under the same name as that used in the original data bank automatically blocks retrieval of the original data bank values—because of MODLER’s data source precedence conventions, as described in Chapter 2. Thus, for example, if during a particular session you wish to create 2 or more single equation models in each case using the same original data values to perform the regressions, you must in between clear the Memoryfile, or at least delete from it the particular affected variables. Otherwise, if your projection period overlaps the regression sample period, you will be in danger of using in the next regression assumed values from a previous projection. If you are making projections beyond the end of the regression sample period, there is in principle no problem, but you still should be aware that in the second case you will be using the Memoryfile as the immediate data source.

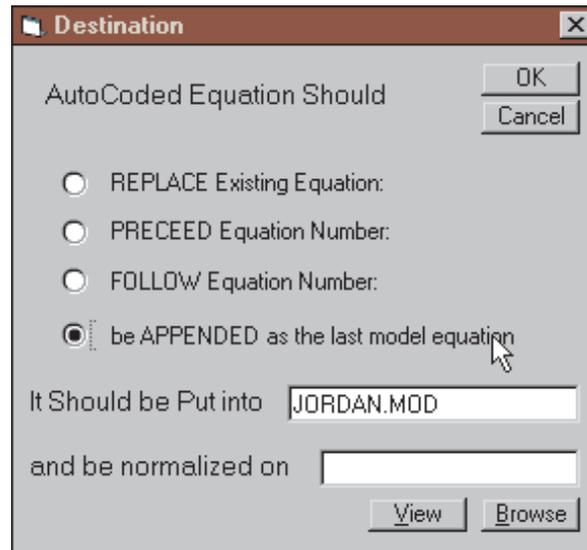
Saving Estimated Equations: Multi-equation Models

In one sense, multi-equation models differ from single equation models only in the fact that the latter generally contain more than a single equation. This statement is both true and false. It is true in the sense that once a single equation model has been created, it can become a multi-equation model by simply adding one or more additional equations. However, Multi-equation models differ from Single Equation models in terms of the operations that *you* must perform. *Any* MODLER-created model can be operated as a multi-equation model, even one containing a single equation, but only Single Equation models can be operated as described in the last section. Multi-equation models are much more flexible in their use, but this flexibility comes at the price of additional complexity; the construction, maintenance, and use of multi-equation models is described in Chapters 6 and 7 of this Guide.

The reason for including this warning here is that in one respect, the process of adding estimated equations to them, multi-equation models involve very little if any additional complexity and only slight differences in comparison with Single Equation models. Once a model exists and has been attached during a MODLER session, the process consists of simply telling the program to add that equation to the model. Replacing an equation involves an essentially similar operation.

One of the effects of creating or attaching any model is to enable the **AutoCode Equation (PutEq)** option of the **Estimate** category of the Parameter Estimation screen menu. Greyed out originally, this option can be clicked on once a model is available for access. Specifically, once a regres-

sion has been performed, simply clicking here causes the results of that regression, the estimated equation, to be included in the model, together with summary statistics and some details of its estimation. The immediate effect is to cause a dialogue box to appear:



Note that the options include both replacing an existing equation and placing a new equation somewhere in the model. The default, as shown, is to append the newly estimated equation, thus adding it as the last equation in the model. The currently attached model is identified, by default, but it is instead possible to include the newly estimated equation in some other existing, but unattached model. Finally, it is also possible to normalize the equation, should this be relevant; as indicated earlier, normalization of equations is considered later.

Corresponding to this menu-based sequence of steps is the command:

PUTEQ

If you press the <Esc> key to return from the Parameter Estimation Screen to the Central Control screen, simply by typing PUTEQ immediately after the regression has been run, the same effect is achieved: the estimated equation is stored with details of its estimation. Additional terms can be added to permit equations to be replaced, the equation to be included in some other model, or other operations performed. This command can also be included in a macro, to permit batched operations. See Chapter 6 for further details.

Sample Period Specification

The regression sample period is specified most easily by clicking on the status bar entries for dates or frequency, as described in Chapter 2. If you click on the status bar of the Central Control screen, the effect is to set, or reset, the dates and frequency globally for the session. If you click on the status bar of the Parameter Estimation screen, the effect is local, generally affecting the regression sample period specifically.

But howsoever the sample period is set, as a standard convention MODLER always attempts to perform a regression over the whole of the specified sample period. Thus if equations involve lags, MODLER attempts to retrieve data for sufficient periods prior to the initial sample period date to allow the regression to be performed as requested. The program truncates the sample period only when sufficient endpoint observations are not available.

In some cases, within the range of available observations, MODLER may discover particular missing observations. In this case, as a default, the program will issue an error message and refuse to execute the regression. At this point, if you wish, you can issue the command:

SET ALLOWNA=ON

which will have the effect of telling the program that missing observations should be ignored. What occurs is that when a within-sample-period missing observation is found for a variable, MODLER will delete the contemporaneous observations on all regression variables, thus in effect reducing the sample period by one observation selectively for each case. In addition, the regression results display produced subsequently will note specifically that missing non-available (NA) observations have been ignored.

There are circumstances in which ignoring missing observations makes sense. In particular, if there are no lagged variables in the regression, and certain other conditions prevail, it can be shown that *ceteris paribus* the effect is only to diminish the efficiency of the parameter estimates. However, in general, omitting NA observations can compromise the properties of a regression. *It is your responsibility to make sure that you have taken into account all the possible effects.*

Trends , Dummy Variables, And Other Such Terms

You can explicitly generate time trends, dummy variables, and other such generated regressor terms if and when you want to include them. For example:

TREND=TIME(1,8401)

Alternatively, the TIME function can be directly embedded in a regression command, either using the ADD option or by typing it into the original command. In general, regression commands can contain both embedded expressions and functions. The most notable exceptions to this rule are the frequency conversion functions, which cannot be included in expressions or in regression commands. A list of MODLER functions is shown on page 70.

However, if you are running regressions in order to build and use models, then you should also take into account that functions defined on a particular time domain (such as TIME, MEAN, SDEV, etc) cannot be included in a model, even if they can be used in regression equations. The use of these functions is not allowed precisely because they can be unintentionally misused in the context of a model; however, as in the case of the variable TREND above, it is always possible to solve the function and then include the variable (or a particular constant) in a model.

Correcting Mistakes

It is easy to make mistakes when specifying functional relationships of any complexity. Even when you build a regression step-by-step using ADD and DROP commands, you are almost certain at some stage to make a typing mistake or to get a series name wrong. When this happens MODLER displays an error message and the offending command, generally offering an opportunity to make the correction.

As a design characteristic, any errors made while using MODLER should be immediately correctable. Occasionally, however, errors can occur that have no obvious source, in which case you may see the message:

Possible Program Malfunction, Please Document and Report

120 Parameter Estimation

This message indicates that the program is “aware” of some sort of problem, but cannot specifically determine its source or, possibly, the appropriate corrective action. Please do document and report such problems.

Functional Forms

The terms in a regression can be names of series or, more generally, functions and expressions. As an example, you can directly estimate a logarithmic relationship by typing

$$\text{LN(IMP)}=\text{f(LN(GNP),LN(CAPI),TREND)}$$

yielding the result:

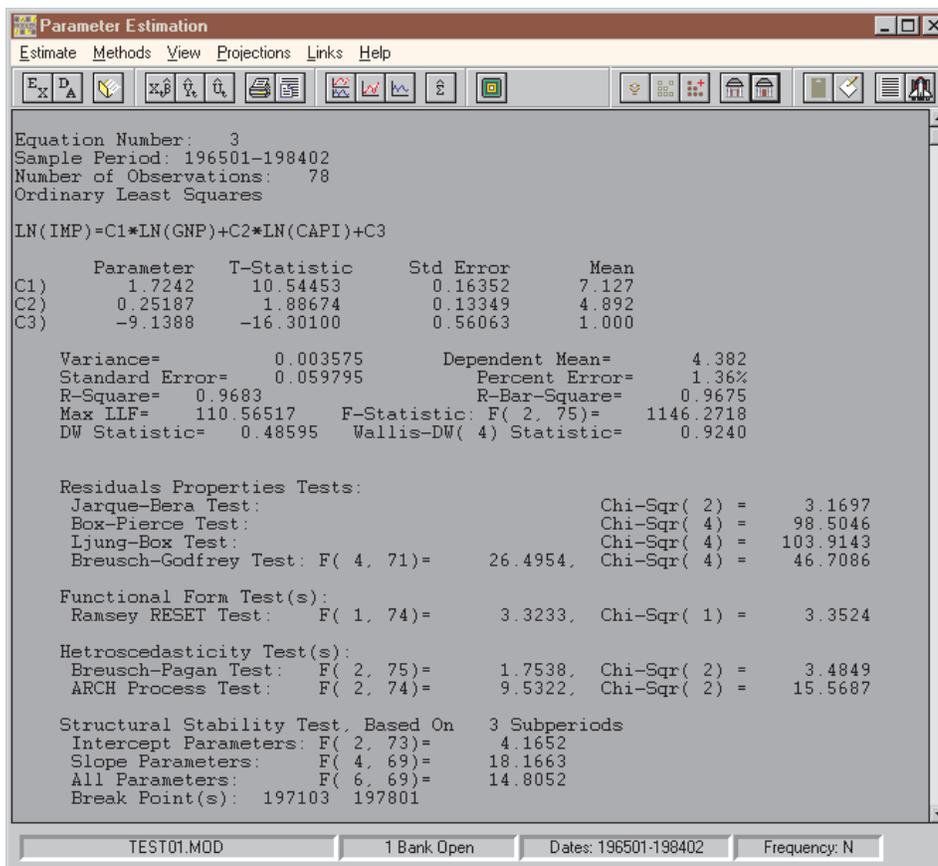


Figure 5-5. Logarithmic Regression

Note that the left-hand-side of a regression may also contain expressions. This flexibility makes it possible to estimate relationships with imposed coefficients, for example:

$$\text{LN(IMP)-0.5*LN(GNP)=F(LN(CAPI),TREND)}$$

Inclusion of a trend in a regression sometimes helps to identify short-term inter-relationships between variables. Another technique for doing this is to express variables as deviations from moving-average values. The following example uses 16-quarter moving averages:

$$\text{IMP-MAVG(16,IMP)=F(GNP-MAVG(16,GNP),CAPI-MAVG(16,CAPI))}$$

Incidentally, observe that this command extends to a second line. If a regression command requires more than one line, you should end each incomplete line with a comma to indicate that further terms will be supplied on the next line. But if you use the comma as a line continuation character, make sure that you break the command after each complete term. *Extraneous comma must not be used: break the expression at existing commas.*

Dynamic Specifications

Lag terms may be specified freely in order to estimate dynamic relationships between variables; you may include as many as 20 terms altogether.

$$\text{IMP=F(GNP,GNP(-1),GNP(-2),CAPI,CAPI(-1),CAPI(-2),IMP(-1),IMP(-2))}$$

But note that it is sometimes better to write a dynamic formulation with difference terms. For example the relationship above is nearly equivalent to:

$$\text{dif(IMP)=f(dif(GNP),lag(1,dif(CAPI))}$$

MODLER provides considerable flexibility in the way in which regression commands may be specified.

Optional Tests: Serial Correlation of the Error Term

One of the assumptions of the general linear model is that equation errors are independently distributed, when in fact it can often be true that they are serially correlated. There are a number of specification tests automatically produced by MODLER that bear on the properties of the error term, as you can see from Figure 5-5. A specific test for autocorrelation is the Durbin-Watson. However, if the regression includes a lagged dependent variable you can also generate the more appropriate Durbin h-statistic, by adding **TEST:DBH** to the regression command:

IMP=F(GNP,CAPL,IMP(-1);TEST:DBH)

But note that, when computing this statistic, the lagged dependent **must** be the last regressor term, as shown here. The Durbin h-statistic is just one of a number of optional tests available; others are described in the online help that can be referenced by preseeing the F1 function key.

A characteristic of tests for autocorrelation, in particular, is that programs such as MODLER also incorporate specific corrections, although you should be aware that it can also be argued that serial correlation can be indicative of more general misspecification errors. You should therefore consider carefully whether you should apply a particular automatic correction or try an alternative specification. In any particular case, MODLER is designed not to force you to take one action rather than another, but instead to support whichever choice you make.

To estimate an equation and include autoregressive correction, requires the addition of **AUTO** to the end of the regression command. For example:

IMP=F(GNP,CAPL;AUTO,ORD=2)

not only specifies autoregressive corrections (Hildreth-Lu, by default), but also defines (ORD=2) that second order autoregressive corrections should be performed; the results are shown at the top of the next page. Incidentally, since the grid search associated with Hildreth-Lu may result in a thousand or more regressions being performed, second order autoregressive equations may take a moment, although in the past few years such “moments” have

shrunk: you are likely to notice a pause only if you are using an older machine, one with a 486 or earlier CPU.

Alternatively, if you specify:

$$\text{IMP}=\text{F}(\text{GNP},\text{CAPI};\text{AUTO}:\text{CO},\text{ORD}=2)$$

the regression will be performed using the Cochrane-Orcutt method. Note the addition of the letters CO following **AUTO**. The results are shown in Figure 5-6.

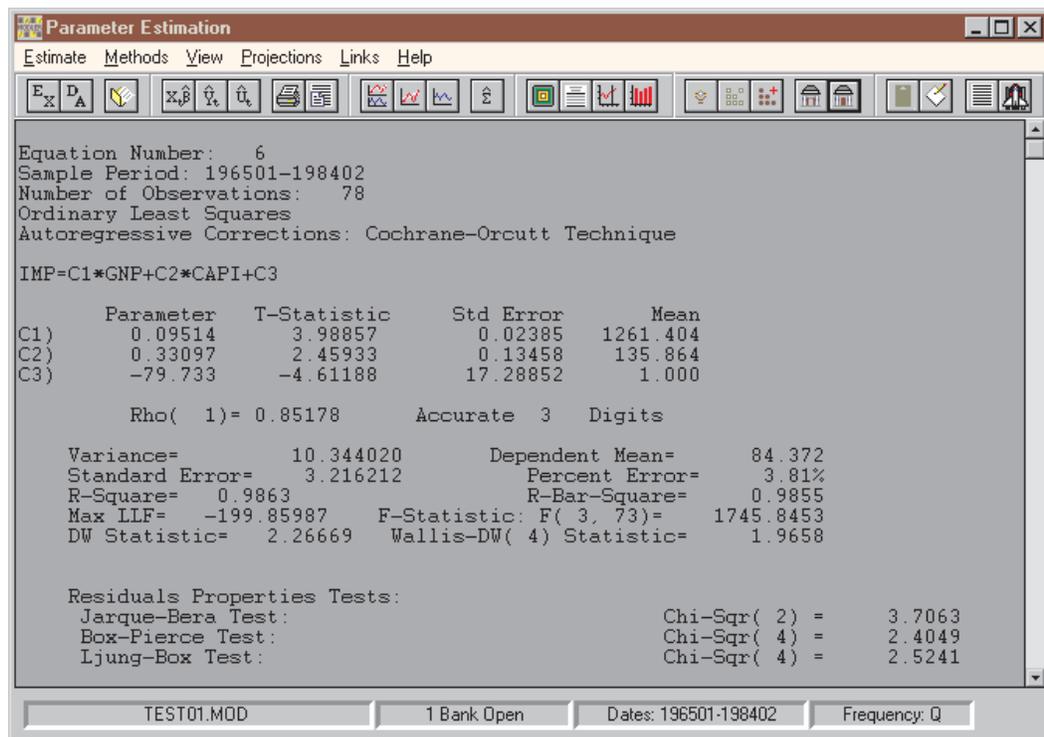


Figure 5-6. Autoregressive Corrections: CO, Order=2

Warning: If you use the AUTO option with others that follow the ; in a regression command, AUTO and its subsidiary commands must be last—otherwise the others will be ignored. For example to combine PRN, and AUTO:CO, type :;PRN,AUTO:CO **not** ;AUTO:CO,PRN.

Distributed Lags

It is often plausible to suppose that variables influence one another with quite long delays. Particularly in the case of quarterly or higher-frequency series this may imply that the effects of changes in an independent variable are distributed over a number of time periods. MODLER provides techniques for estimating patterns of distributed lag coefficients represented by polynomial curves.

As an example, consider the possible influence of import prices on the volume of imports. When the prices of imports are high relative to those of domestic products, the volume of imports is likely to be depressed as purchasers switch to domestic sources of supply. However, the influence of changes in relative import prices is not immediate, in part because of the existence of contracts; if they make themselves felt, these changes may cumulate over several quarters or even over years.

To calculate an index of the relative price of imports using the data in DEMOBANK, specify:

$$\begin{aligned} \text{PDIMP} &= \text{IMP\$} / \text{IMP} \\ \text{PDGNP} &= \text{GNP\$} / \text{GNP} \\ \text{RPIMP} &= \text{PDIMP} / \text{PDGNP} \end{aligned}$$

An unlagged regression for imports as a function of GNP and relative prices can then be specified as:

$$\text{LN}(\text{IMP}) = \text{F}(\text{LN}(\text{GNP}), \text{LN}(\text{RPIMP}))$$

In contrast, to define a distributed lag, it is necessary to state the total number of lag terms, L , and the degree of the polynomial approximation, D . Suppose that we expect relative import prices to continue having some effect for up to three years ($L=12$) and want to allow at least some curvature in the pattern of coefficients ($D=2$). An appropriate regression command might be:

$$\text{LN}(\text{IMP}) = \text{F}(\text{LN}(\text{GNP}), \text{LN}(\text{RPIMP}) \langle \text{D}=2, \text{L}=12 \rangle)$$

which yields the results shown in Figure 5-7.

Equation Number: 6				
Sample Period: 196801-198402				
Number of Observations: 66				
Ordinary Least Squares				
Polynomial Distributed Lag				
LN(IMP)=C1*LN(GNP)+L1*LN(IMP\$/IMP\$/GNP\$/GNP)<D=2,L=12>+C2				
	Parameter	T-Statistic	Std Error	Mean
C1)	1.4620	23.79197	0.06145	7.172
C2)	-12.971	-23.30955	0.55649	1.000
Variance=		0.001383	Dependent Mean=	4.480
Standard Error=		0.037193	Percent Error=	0.83%
R-Square=		0.9801	R-Bar-Square=	0.9788
Max LLF=		126.19807	F-Statistic: F(4, 61)=	752.8839
Polynomial Weights: L1				
Degree of Polynomial= 2				
No Restrictions on Roots				
Estimated Polynomial Coefficients - f(Z):				
Term	Coeff	T-Statistic		
Z**0	-0.14988	-6.87334		
Z**1	0.03896	3.96828		
Z**2	-0.00244	-3.01897		
Lag Weights:				
Period	Weight	T-Statistic		
T- 0	-0.14988	-6.87334	*	
T- 1	-0.11335	-8.29872	.	*
T- 2	-0.08171	-9.91846	.	*
T- 3	-0.05493	-7.86967	.	*
T- 4	-0.03304	-3.83545	.	*
T- 5	-0.01602	-1.56368	.	*
T- 6	-0.00388	-0.35810	.	*
T- 7	0.00338	0.33279	.	*
T- 8	0.00577	0.68611	.	*
T- 9	0.00328	0.50231	.	*
T-10	-0.00408	-0.53475	.	*
T-11	-0.01632	-1.24475	.	*
T-12	-0.03344	-1.56924	.	*
Sum	-0.49422	-11.11626	Mean Lag	2.43
			Median Lag	1.86
DW Statistic=		1.33251	Wallis-DW(4) Statistic=	2.3588
Residuals Properties Tests:				
Jarque-Bera Test:		Chi-Sqr(2) =		6.9496
Box-Pierce Test:		Chi-Sqr(4) =		10.4920
Ljung-Box Test:		Chi-Sqr(4) =		11.1402
Breusch-Godfrey Test: F(4, 57)=	2.7074,	Chi-Sqr(4) =		10.5374
Functional Form Test(s):				
Ramsey RESET Test: F(1, 60)=	8.2048,	Chi-Sqr(1) =		7.9396
Heteroscedasticity Test(s):				
Breusch-Pagan Test: F(4, 61)=	1.3885,	Chi-Sqr(4) =		5.5076
ARCH Process Test: F(2, 62)=	0.3198,	Chi-Sqr(2) =		0.6535
Structural Stability Test, Based On 3 Subperiods				
Intercept Parameters: F(2, 59)=		1.3711		
Slope Parameters: F(8, 51)=		2.2398		
All Parameters: F(10, 51)=		2.1121		
Break Point(s):	197303	197901		

Figure 5-7. Distributed Lags Example

However, MODLER also allows you to insist that the coefficients start from zero, NEAR, or end at zero, FAR, or BOTH. It may be reasonable instead to specify “FAR” restrictions:

$$\text{LN(IMP)=F(LN(GNP),LN(RPIMP)<D=2,L=12,FAR>)}$$

Additional Regression Facilities

The focus so far in this chapter has been possible variations on the basic regression command. Almost without exception, the techniques considered can be interpreted as simply involving variable transformations and under this interpretation the discussion arguably has been limited essentially to Ordinary Least Squares. Five additional, frequently-used, MODLER regression options are:

- (i) Restricted Least Squares
- (ii) Weighted Least Squares
- (iii) Ordinary Ridge Regression
- (iv) Logistic Curve
- (v) Two Stage Least Squares

The commands to execute each of these still involve only slight modifications of the basic regression command.

The stress, as before, will be the commands themselves. The reason is that equations are the stuff of econometric models and therefore their specification must be understood. However, MODLER also provides additional support via menu options. Observe that clicking on the Methods menu category of the Parameter Estimation screen displays a choice of techniques, as partially shown in Figure 5-8. Furthermore, given the basic equation specification, making choices among these techniques has the immediate effect of automatically modifying MODLER regression command(s). Thus it is not necessary in all cases to modify these commands manually when switching between techniques.

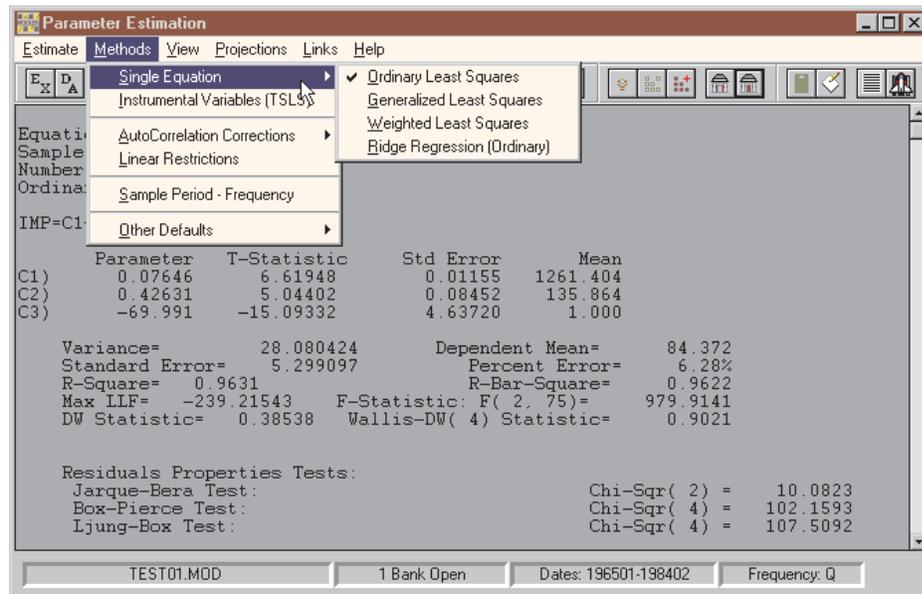


Figure 5-8. Regression Methods and Options

Restricted Least Squares

In order to perform Restricted Least Squares, you must begin by specifying the restrictions. The command is:

RESTRICTIONS:R=r₁₁,r₁₂,...,r_{1k};r₂₁,...,r_{2k};r_{n1} ,...,r_{nk};C=c₁,...,c_n

where r_{ij} denotes values in the restrictions matrix, R , and the c_i are elements of the C vector, using the Theil-Johnston notation. Note that semi-colons separate both rows of the R matrix and the elements of the R matrix from those of the C vector. Up to 15 restriction rows may be specified per regression—these must be independent linear restrictions on the parameters.

Once the RESTRICTIONS command is issued, the characteristics of the restrictions have been specified, but to actually perform a Restricted Least Squares regression, you must also insert ;RB=C into the regression command; for example:

$$Y=F(X_1,X_2,\dots,X_{K-1};RB=C)$$

Regressions performed without this insert will ignore the restrictions.

Weighted Least Squares

The command to perform Weighted Least Squares also involves the insertion of a term: WLS:Z, where Z is the name of the weight series. For example:

$$Y=F(X_1,X_2,\dots,X_{K-1};WLS:Z)$$

The observations on Z will be divided into the observations on all the regression terms, making this command in the simple case equivalent to:

$$Y/Z=F(X_1/Z,X_2/Z,\dots,X_{K-1}/Z,1/Z;NOCON)$$

In order to perform Weighted Least Squares where each term in the regression is weighted by a separate variable, you simply need to divide each term by the appropriate weight variable. But bear in mind that in the simple case, if $Z_i, i=1,2,\dots,T$ denotes the time series elements of Z, the use of Weighted Least Squares implies the hypothesis that the covariance matrix of the estimator, V, takes the form:

$$V = \text{diag} (s_1^2, \dots, s_T^2)$$

In the general case, the elements of V are not so easily specified.

Ordinary Ridge Regression

The Ridge regression command involves the specification of the Ridge Trace constant, c:

$$Y=F(X_1,X_2,\dots,X_{K-1};RIDGE=c)$$

where c is a real value scalar number.

Logistic Curve Model

The Logistic Curve Model can be defined in terms of MODLER's LGT function:

$$\text{LGT}(Y)=F(X_1,X_2,\dots,X_{K-1})$$

where:

$$\text{LGT}(Y) = \ln \left(\frac{Y}{1 - Y} \right)$$

defines the LGT function; **Y must** take a value between 0 and 1. The Logistic Model can be used in conjunction with Weighted Least Squares, in order to provide a GLS estimator. The argument term of the LGT function can be a single variable or an expression. Estimated equations can be normalized in terms of the variables that appear in the argument provided that the expression in the argument is an analytic function, and thus that an inverse function exists.

Two Stage Least Squares

When an equation in a simultaneous model contains one or more (unlagged) model endogenous variables among its regressors, those regressors can be expected to be correlated with the equation disturbances. In this case, OLS will produce inconsistent equation parameter estimates. Two Stage Least Squares is a method that works by effectively "purging" the right-hand-side endogenous variables of that correlation in the first stage so that a second stage pass which applies OLS to transformed observations will yield consistent parameter estimators.

There are various ways in which to interpret the "purging" process, but in practice an essential characteristic of 2SLS is that exogenous (instrument) variables must be chosen to act as first stage regressors. All other things equal, these exogenous variables should have two properties in particular. First, they should be closely related to the right-hand-side endogenous

variables. Second, they should be statistically independent of the structural equation errors. Beyond this, the precise choice of variables generally depends upon the specification of the relationships that define the model. If you have any question about the nature of the 2SLS method, you should consult an econometrics textbook for a complete description of the method and its motivation.

Instruments Command

In order to perform Two Stage Least Squares using MODLER it is first necessary to specify the first stage instruments using the INSTRUMENTS command:

INSTRUMENTS:Z₁,Z₂,...,Z_k

where the variables Z_1, Z_2, \dots, Z_k are the instruments. These should all be accessible at the time the INSTRUMENTS command is executed, since MODLER will immediately attempt to retrieve them once the Enter key is pressed. Furthermore, they must all be named series: any variables that involve lags or other transformations, functions, or expressions must first be evaluated and given a series name; the variables can be stored either in the Memory File or a Data Bank. As many as 40 instruments may be used at a time; they must all have a common date range—moreover, the same date range must be used for the subsequent, second stage regression(s).

In general, the instruments might be expected to consist of the entire set of model exogenous variables. However, MODLER does not in any way attempt to impose this requirement. For example, the instruments used do not need to include the ostensibly predetermined variables that appear in the second stage regression(s).

Second Stage Regression Command

The Two Stage Least Squares regression command begins with the prefix, TSLS(m):

TSLS(m):y=F(x₁,...,x_m,x_{m+1}...,x_n)

where m refers to the number of regressor terms that are considered endogenous, among the n regressor terms that appear in the command. The

endogenous regressors must be located, in a group, immediately after F(. They must precede any regressor terms considered to be predetermined. There is otherwise no way for MODLER to determine which variables the user considers to be right-hand-side endogenous.

The Two Stage Least Squares regression command may be generalized to allow for autoregressive corrections and distributed lags (involving either right-hand-side endogenous or included predetermined variables). Similarly, the second stage constant term may be included or omitted. The generalizations are made in an exactly analogous fashion as in the OLS case. For example, the command:

TLS(m):y=F(x₁,...,x_m,x_{m+1}...,x_n;AUTO)

specifies that Two Stage Least Squares and autoregressive corrections are to be performed, with the autoregressive corrections made using the Hildreth-Lu method.

It should also be noted that Two Stage Least Squares regressions can be performed either as a batched group of regressions or mixed with OLS (and other) regressions. The TLS prefix identifies equations that are to be estimated by Two Stage Least Squares and if this prefix is absent, MODLER will not perform a 2SLS regression. Of course, for Two Stage Least Squares to be performed at all the INSTRUMENTS command must have been executed previously during the session.

All the supplementary single equation estimator commands, such as the PRINT GRAPH and SAVE RESIDUALS commands, can be used with Two Stage Least Squares regressions. Whenever necessary, MODLER will automatically allow for any differences related to the particular regression method used.

Additional regression options and further details concerning the estimators described above can be found by clicking on the Help menu category of the Parameter Estimation Screen, or while that screen is displayed by pressing the <F1> function key.

