Chapter 19
The SYSLIN Procedure

Chapter Table of Contents

Chapter 19
The SYSLIN Procedure

Overview

The SYSLIN procedure estimates parameters in an interdependent system of linear regression equations.

Ordinary least squares (OLS) estimates are biased and inconsistent when current period endogenous variables appear as regressors in other equations in the system. The errors of a set of related regression equations are often correlated, and the efficiency of the estimates can be improved by taking these correlations into account. The SYSLIN procedure provides several techniques which produce consistent and asymptotically efficient estimates for systems of regression equations.

The SYSLIN procedure provides the following estimation methods:

- ordinary least squares (OLS)
- two-stage least squares (2SLS)
- limited information maximum likelihood (LIML)
- K-class
- seemingly unrelated regressions (SUR)
- iterated seemingly unrelated regressions (ITSUR)
- three-stage least squares (3SLS)
- iterated three-stage least squares (IT3SLS)
- full information maximum likelihood (FIML)
- minimum expected loss (MELO)

Other features of the SYSLIN procedure enable you to:

- impose linear restrictions on the parameter estimates.
- test linear hypotheses about the parameters.
- write predicted and residual values to an output SAS data set.
- write parameter estimates to an output SAS data set.
- write the crossproducts matrix (SSCP) to an output SAS data set.
- use raw data, correlations, covariances, or cross products as input.

Getting Started

This section introduces the use of the SYSLIN procedure. The problem of dependent regressors is introduced using a supply-demand example. This section explains the terminology used for variables in a system of regression equations and introduces the SYSLIN procedure statements for declaring the roles the variables play. The syntax used for the different estimation methods and the output produced is shown.

An Example Model

In simultaneous systems of equations, endogenous variables are determined jointly rather than sequentially. Consider the following demand and supply functions for some product:

$$Q_D = a_1 + b_1 P + c_1 Y + d_1 S + \epsilon_1 \qquad \text{(demand)}$$

$$Q_S = a_2 + b_2 P + c_2 U + \epsilon_2 \qquad \text{(supply)}$$

$$Q = Q_D = Q_S \qquad \text{(market equilibrium)}$$

The variables in this system are as follows:

Q_D	quantity demanded
Q_S	quantity supplied
Q	the observed quantity sold, which equates quantity supplied and quantity demanded in equilibrium
P	price per unit
Y	income
S	price of substitutes
U	unit cost
ϵ_1	the random error term for the demand equation
ϵ_2	the random error term for the supply equation

In this system, quantity demanded depends on price, income, and the price of substitutes. Consumers normally purchase more of a product when prices are lower and when income and the price of substitute goods are higher. Quantity supplied depends on price and the unit cost of production. Producers will supply more when price is high and when unit cost is low. The actual price and quantity sold are determined jointly by the values that equate demand and supply.

Since price and quantity are jointly endogenous variables, both structural equations are necessary to adequately describe the observed values. A critical assumption of OLS is that the regressors are uncorrelated with the residual. When current endogenous variables appear as regressors in other equations (endogenous variables depend on each other), this assumption is violated and the OLS parameter estimates are biased and inconsistent. The bias caused by the violated assumptions is called *Simultaneous equation bias*. Neither the demand nor supply equation can be estimated consistently by OLS.

Variables in a System of Equations

Before explaining how to use the SYSLIN procedure, it is useful to define some terms. The variables in a system of equations can be classified as follows:

- *Endogenous variables*, which are also called *jointly dependent* or *response variables*, are the variables determined by the system. Endogenous variables can also appear on the right-hand side of equations.
- *Exogenous variables* are independent variables that do not depend on any of the endogenous variables in the system.
- *Predetermined variables* include both the exogenous variables and *lagged endogenous variables*, which are past values of endogenous variables determined at previous time periods. PROC SYSLIN does not compute lagged values; any lagged endogenous variables must be computed in a preceding DATA step.
- *Instrumental variables* are predetermined variables used in obtaining predicted values for the current period endogenous variables by a first-stage regression. The use of instrumental variables characterizes estimation methods such as two-stage least squares and three-stage least squares. Instrumental variables estimation methods substitute these first-stage predicted values for endogenous variables when they appear as regressors in model equations.

Using PROC SYSLIN

First specify the input data set and estimation method on the PROC SYSLIN statement. If any model uses dependent regressors, and you are using an instrumental variables regression method, declare the dependent regressors with an ENDOGENOUS statement and declare the instruments with an INSTRUMENTS statement. Next, use MODEL statements to specify the structural equations of the system.

The use of different estimation methods is shown by the following examples. These examples use simulated data (not shown).

OLS Estimation

PROC SYSLIN performs OLS regression if you do not specify a method of estimation in the PROC SYSLIN statement. OLS does not use instruments, so the ENDOGENOUS and INSTRUMENTS statements can be omitted.

The following statements estimate the supply and demand model shown previously:

```
proc syslin data=in;
   demand: model q = p y s;
   supply: model q = p u;
run;
```

The PROC SYSLIN output for the demand equation is shown in Figure 19.1, and the output for the supply equation is shown in Figure 19.2.

```
                         The SYSLIN Procedure
                   Ordinary Least Squares Estimation

                     Model                  DEMAND
                     Dependent Variable          q
                     Label                Quantity


                          Analysis of Variance

                                   Sum of        Mean
     Source                 DF    Squares      Square    F Value    Pr > F

     Model                   3   9.587891    3.195964     398.31    <.0001
     Error                  56   0.449336    0.008024
     Corrected Total        59   10.03723


             Root MSE           0.08958    R-Square       0.95523
             Dependent Mean     1.30095    Adj R-Sq       0.95283
             Coeff Var          6.88541


                          Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  -0.47677 0.210239   -2.27   0.0272 Intercept
p              1  0.123324 0.105177    1.17   0.2459 Price
y              1  0.201282 0.032403    6.21   <.0001 Income
s              1  0.167258 0.024091    6.94   <.0001 Price of Substitutes
```

Figure 19.1. OLS Results for Demand Equation

```
                        The SYSLIN Procedure
                 Ordinary Least Squares Estimation

                   Model                   SUPPLY
                   Dependent Variable           q
                   Label                 Quantity


                        Analysis of Variance

                                    Sum of        Mean
     Source                 DF     Squares      Square    F Value    Pr > F

     Model                   2    9.033890    4.516945     256.61    <.0001
     Error                  57    1.003337    0.017602
     Corrected Total        59    10.03723


             Root MSE             0.13267    R-Square       0.90004
             Dependent Mean       1.30095    Adj R-Sq       0.89653
             Coeff Var           10.19821


                        Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  -0.30390 0.471397   -0.64   0.5217 Intercept
p              1  1.218743 0.053914   22.61   <.0001 Price
u              1  -1.07757 0.234150   -4.60   <.0001 Unit Cost
```

Figure 19.2. OLS Results for Supply Equation

For each MODEL statement, the output first shows the model label and dependent variable name and label. This is followed by an Analysis of Variance table for the model, which shows the model, error, and total mean squares, and an *F* test for the no-regression hypothesis. Next, the procedure prints the root mean square error, dependent variable mean and coefficient of variation, and the R^2 and adjusted R^2 statistics.

Finally, the table of parameter estimates shows the estimated regression coefficients, standard errors, and *t*-tests. You would expect the price coefficient in a demand equation to be negative. However, note that the OLS estimate of the price coefficient P in the demand equation (.1233) has a positive sign. This could be caused by simultaneous equation bias.

Two-Stage Least Squares Estimation

In the supply and demand model, P is an endogenous variable, and consequently the OLS estimates are biased. The following example estimates this model using two-stage least squares.

```
proc syslin data=in 2sls;
   endogenous  p;
   instruments y u s;
   demand: model q = p y s;
   supply: model q = p u;
run;
```

The 2SLS option on the PROC SYSLIN statement specifies the two-stage least-squares method. The ENDOGENOUS statement specifies that P is an endogenous regressor for which first-stage predicted values are substituted. You only need to declare an endogenous variable in the ENDOGENOUS statement if it is used as a regressor; thus although Q is endogenous in this model, it is not necessary to list it in the ENDOGENOUS statement.

Usually, all predetermined variables that appear in the system are used as instruments. The INSTRUMENTS statement specifies that the exogenous variables Y, U, and S are used as instruments for the first-stage regression to predict P.

The 2SLS results are shown in Figure 19.3 and Figure 19.4. The first-stage regressions are not shown. To see the first-stage regression results, use the FIRST option on the MODEL statement.

```
                              The SYSLIN Procedure
                       Two-Stage Least Squares Estimation

                         Model                   DEMAND
                         Dependent Variable           q
                         Label                 Quantity


                              Analysis of Variance

                                     Sum of        Mean
     Source                 DF      Squares      Square    F Value    Pr > F

     Model                   3     9.670882    3.223627     115.58    <.0001
     Error                  56     1.561944    0.027892
     Corrected Total        59     10.03723


             Root MSE             0.16701    R-Square       0.86095
             Dependent Mean       1.30095    Adj R-Sq       0.85350
             Coeff Var           12.83740


                              Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  1.901040 1.171224    1.62   0.1102 Intercept
p              1  -1.11518 0.607391   -1.84   0.0717 Price
y              1  0.419544 0.117954    3.56   0.0008 Income
s              1  0.331475 0.088472    3.75   0.0004 Price of Substitutes
```

Figure 19.3. 2SLS Results for Demand Equation

```
                         The SYSLIN Procedure
                  Two-Stage Least Squares Estimation

                    Model                   SUPPLY
                    Dependent Variable           q
                    Label                 Quantity


                         Analysis of Variance

                               Sum of        Mean
     Source               DF     Squares      Square    F Value    Pr > F

     Model                 2    9.646098    4.823049     253.96    <.0001
     Error                57    1.082503    0.018991
     Corrected Total      59    10.03723


             Root MSE            0.13781    R-Square       0.89910
             Dependent Mean      1.30095    Adj R-Sq       0.89556
             Coeff Var          10.59291


                         Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  -0.51878 0.490999   -1.06   0.2952 Intercept
p              1  1.333080 0.059271   22.49   <.0001 Price
u              1  -1.14623 0.243491   -4.71   <.0001 Unit Cost
```

Figure 19.4. 2SLS Results for Supply Equation

The 2SLS output is similar in form to the OLS output. However, the 2SLS results are based on predicted values for the endogenous regressors from the first stage instrumental regressions. This makes the analysis of variance table and the R^2 statistics difficult to interpret. See the sections "ANOVA Table for Instrumental Variables Methods" and "The R^2 Statistics" later in this chapter for details.

Note that, unlike the OLS results, the 2SLS estimate for the P coefficient in the demand equation (-1.115) is negative.

LIML, K-Class, and MELO Estimation

To obtain limited information maximum likelihood, general K-class, or minimum expected loss estimates, use the ENDOGENOUS, INSTRUMENTS, and MODEL statements as in the 2SLS case but specify the LIML, K=, or MELO option instead of 2SLS in the PROC SYSLIN statement. The following statements show this for K-class estimation.

```
proc syslin data=in k=.5;
   endogenous  p;
   instruments y u s;
   demand: model q = p y s;
   supply: model q = p u;
run;
```

For more information on these estimation methods see the "Estimation Methods" in the "Details" section and consult econometrics textbooks.

SUR, 3SLS, and FIML Estimation

In a multivariate regression model, the errors in different equations may be correlated. In this case the efficiency of the estimation may be improved by taking these cross-equation correlations into account.

Seemingly Unrelated Regression

Seemingly unrelated regression (SUR), also called joint generalized least squares (JGLS) or Zellner estimation, is a generalization of OLS for multi-equation systems. Like OLS, the SUR method assumes that all the regressors are independent variables, but SUR uses the correlations among the errors in different equations to improve the regression estimates. The SUR method requires an initial OLS regression to compute residuals. The OLS residuals are used to estimate the cross-equation covariance matrix.

The SUR option on the PROC SYSLIN statement specifies seemingly unrelated regression, as shown in the following statements:

```
proc syslin data=in sur;
   demand: model q = p y s;
   supply: model q = p u;
run;
```

INSTRUMENTS and ENDOGENOUS statements are not needed for SUR, since the SUR method assumes there are no endogenous regressors. For SUR to be effective, the models must use different regressors. SUR produces the same results as OLS unless the model contains at least one regressor not used in the other equations.

Three-Stage Least Squares

The three-stage least-squares method generalizes the two-stage least-squares method to take account of the correlations between equations in the same way that SUR generalizes OLS. Three-stage least squares requires three steps: first-stage regressions to get predicted values for the endogenous regressors; a two-stage least-squares step to get residuals to estimate the cross-equation correlation matrix; and the final 3SLS estimation step.

The 3SLS option on the PROC SYSLIN statement specifies the three-stage least-squares method, as shown in the following statements.

```
proc syslin data=in 3sls;
   endogenous  p;
   instruments y u s;
   demand: model q = p y s;
   supply: model q = p u;
run;
```

The 3SLS output begins with a two-stage least-squares regression to estimate the cross-model correlation matrix. This output is the same as the 2SLS results shown in Figure 19.3 and Figure 19.4, and is not repeated here. The next part of the 3SLS output prints the cross-model correlation matrix computed from the 2SLS residuals. This output is shown in Figure 19.5 and includes the cross-model covariances, correlations, the inverse of the correlation matrix, and the inverse covariance matrix.

```
                    The SYSLIN Procedure
            Three-Stage Least Squares Estimation

                   Cross Model Covariance

                          DEMAND        SUPPLY

            DEMAND        0.027892      -.011283
            SUPPLY        -.011283      0.018991


                  Cross Model Correlation

                          DEMAND        SUPPLY

            DEMAND         1.00000      -0.49022
            SUPPLY        -0.49022       1.00000


              Cross Model Inverse Correlation

                          DEMAND        SUPPLY

            DEMAND         1.31634       0.64530
            SUPPLY         0.64530       1.31634


               Cross Model Inverse Covariance

                          DEMAND        SUPPLY

            DEMAND         47.1945       28.0380
            SUPPLY         28.0380       69.3130
```

Figure 19.5. Estimated Cross-Model Covariances used for 3SLS Estimates

The final 3SLS estimates are shown in Figure 19.6.

```
                               The SYSLIN Procedure
                       Three-Stage Least Squares Estimation

                      System Weighted MSE            0.5711
                      Degrees of freedom                113
                      System Weighted R-Square       0.9627


                          Model                   DEMAND
                          Dependent Variable           q
                          Label                 Quantity


                               Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  1.980261 1.169169    1.69   0.0959 Intercept
p              1  -1.17654 0.605012   -1.94   0.0568 Price
y              1  0.404115 0.117179    3.45   0.0011 Income
s              1  0.359204 0.085077    4.22   <.0001 Price of Substitutes


                          Model                   SUPPLY
                          Dependent Variable           q
                          Label                 Quantity


                               Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  -0.51878 0.490999   -1.06   0.2952 Intercept
p              1  1.333080 0.059271   22.49   <.0001 Price
u              1  -1.14623 0.243491   -4.71   <.0001 Unit Cost
```

Figure 19.6. Three-Stage Least Squares Results

This output first prints the system weighted mean square error and system weighted R^2 statistics. The system weighted MSE and system weighted R^2 measure the fit of the joint model obtained by stacking all the models together and performing a single regression with the stacked observations weighted by the inverse of the model error variances. See the section "The R^2 Statistics" for details.

Next, the table of 3SLS parameter estimates for each model is printed. This output has the same form as for the other estimation methods.

Note that the 3SLS and 2SLS results may be the same in some cases. This results from the same principle that causes OLS and SUR results to be identical unless an equation includes a regressor not used in the other equations of the system. However, the application of this principle is more complex when instrumental variables are used. When all the exogenous variables are used as instruments, linear combinations of all the exogenous variables appear in the third-stage regressions through substitution of first-stage predicted values.

In this example, 3SLS produces different (and, it is hoped, more efficient) estimates for the demand equation. However, the 3SLS and 2SLS results for the supply equation are the same. This is because the supply equation has one endogenous regressor and one exogenous regressor not used in other equations. In contrast, the demand equation has fewer endogenous regressors than exogenous regressors not used in other equations in the system.

Full Information Maximum Likelihood

The FIML option on the PROC SYSLIN statement specifies the full information maximum likelihood method, as shown in the following statements.

```
proc syslin data=in fiml;
   endogenous  p q;
   instruments y u s;
   demand: model q = p y s;
   supply: model q = p u;
run;
```

The FIML results are shown in Figure 19.7.

```
                              The SYSLIN Procedure
                 Full-Information Maximum Likelihood Estimation

        NOTE: Convergence criterion met at iteration 3.


                         Model                   DEMAND
                         Dependent Variable           q
                         Label                 Quantity


                              Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  1.988529 1.233625    1.61   0.1126 Intercept
p              1  -1.18147 0.652274   -1.81   0.0755 Price
y              1  0.402310 0.107269    3.75   0.0004 Income
s              1  0.361345 0.103816    3.48   0.0010 Price of Substitutes


                         Model                   SUPPLY
                         Dependent Variable           q
                         Label                 Quantity


                              Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  -0.52443 0.479522   -1.09   0.2787 Intercept
p              1  1.336083 0.057939   23.06   <.0001 Price
u              1  -1.14804 0.237793   -4.83   <.0001 Unit Cost
```

Figure 19.7. FIML Results

Computing Reduced Form Estimates

A system of structural equations with endogenous regressors can be represented as functions only of the predetermined variables. For this to be possible, there must be as many equations as endogenous variables. If there are more endogenous variables than regression models, you can use IDENTITY statements to complete the system. See "Reduced Form Estimates" in the "Computational Details" section later in this chapter for details.

The REDUCED option on the PROC SYSLIN statement prints reduced form estimates. The following statements show this using the 3SLS estimates of the structural parameters.

```
proc syslin data=in 3sls reduced;
   endogenous  p;
   instruments y u s;
   demand: model q = p y s;
   supply: model q = p u;
run;
```

The first four pages of this output were as shown previously and are not repeated here. (See Figure 19.3, Figure 19.4, Figure 19.5, and Figure 19.6.) The final page of the output from this example contains the reduced form coefficients from the 3SLS structural estimates, as shown in Figure 19.8.

```
                        The SYSLIN Procedure
                Three-Stage Least Squares Estimation

                        Endogenous Variables

                                 p            q

                DEMAND    1.176539            1
                SUPPLY    -1.33308            1


                        Exogenous Variables

               Intercept           y            s            u

DEMAND          1.980261    0.404115     0.359204            0
SUPPLY          -0.51878           0            0     -1.14623


                   Inverse Endogenous Variables

                           DEMAND       SUPPLY

                   p     0.398467     -0.39847
                   q     0.531188     0.468812


                           Reduced Form

           Intercept           y            s            u

    p       0.995786    0.161027     0.143131     0.456736
    q        0.80868    0.214661     0.190805     -0.53737
```

Figure 19.8. Reduced Form 3SLS Results

Restricting Parameter Estimates

You can impose restrictions on the parameter estimates with RESTRICT and SRESTRICT statements. The RESTRICT statement imposes linear restrictions on parameters in the equation specified by the preceding MODEL statement. The SRESTRICT statement imposes linear restrictions that relate parameters in different models.

To impose restrictions involving parameters in different equations, use the SRESTRICT statement. Specify the parameters in the linear hypothesis as *model-label.regressor-name*. (If the MODEL statement does not have a label, you can use the dependent variable name as the label for the model, provided the dependent variable uniquely labels the model.)

Tests for the significance of the restrictions are printed when RESTRICT or SRESTRICT statements are used. You can label RESTRICT and SRESTRICT statements to identify the restrictions in the output.

The RESTRICT statement in the following example restricts the price coefficient in the demand equation to equal .015. The SRESTRICT statement restricts the estimate of the income coefficient in the demand equation to be .01 times the estimate of the unit cost coefficient in the supply equation.

```
proc syslin data=in 3sls;
   endogenous  p;
   instruments y u s;
   demand: model q = p y s;
   peq015: restrict p = .015;
   supply: model q = p u;
   yeq01u: srestrict demand.y = .01 * supply.u;
run;
```

The restricted estimation results are shown in Figure 19.9.

```
                              The SYSLIN Procedure
                       Three-Stage Least Squares Estimation

                         Model                   DEMAND
                         Dependent Variable           q
                         Label                 Quantity


                              Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  -0.46584 0.053307   -8.74   <.0001 Intercept
p              1  0.015000        0     .       .    Price
y              1  -0.00679 0.002357   -2.88   0.0056 Income
s              1  0.325589 0.009872   32.98   <.0001 Price of Substitutes
RESTRICT      -1  50.59341 7.464990    6.78   <.0001 PEQ015


                         Model                   SUPPLY
                         Dependent Variable           q
                         Label                 Quantity


                              Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  -1.31894 0.477633   -2.76   0.0077 Intercept
p              1  1.291718 0.059101   21.86   <.0001 Price
u              1  -0.67887 0.235679   -2.88   0.0056 Unit Cost


                              Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

RESTRICT      -1  342.3611 38.12103    8.98   <.0001 YEQ01U
```

Figure 19.9. Restricted Estimates

The standard error for P in the demand equation is 0, since the value of the P coefficient was specified by the RESTRICT statement and not estimated from the data. The Parameter Estimates table for the demand equation contains an additional row for the restriction specified by the RESTRICT statement. The "parameter estimate" for the restriction is the value of the Lagrange multiplier used to impose the restriction. The restriction is highly "significant" ($t = 6.777$), which means that the data are not consistent with the restriction, and the model does not fit as well with the restriction imposed. See the section "RESTRICT Statement" for more information.

After the Parameter Estimates table for the supply equation, the results for the cross model restrictions are printed. This shows that the restriction specified by the SRESTRICT statement is not consistent with the data ($t = 8.98$). See the section "SRESTRICT Statement" for more information.

Testing Parameters

You can test linear hypotheses about the model parameters with TEST and STEST statements. The TEST statement tests hypotheses about parameters in the equation specified by the preceding MODEL statement. The STEST statement tests hypotheses that relate parameters in different models.

For example, the following statements test the hypothesis that the price coefficient in the demand equation is equal to .015.

```
proc syslin data=in 3sls;
   endogenous  p;
   instruments y u s;
   demand: model q = p y s;
   test_1: test p = .015;
   supply: model q = p u;
run;
```

The TEST statement results are shown in Figure 19.10. This reports an *F*-test for the hypothesis specified by the TEST statement. In this case the *F* statistic is 6.79 (3.879/.571) with 1 and 113 degrees of freedom. The *p*-value for this *F* statistic is .0104, which indicates that the hypothesis tested is almost but not quite rejected at the .01 level. See the section "TEST Statement" for more information.

```
                              The SYSLIN Procedure
                      Three-Stage Least Squares Estimation

                     System Weighted MSE            0.5711
                     Degrees of freedom                113
                     System Weighted R-Square       0.9627


                         Model                    DEMAND
                         Dependent Variable            q
                         Label                  Quantity


                              Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  1.980261 1.169169    1.69   0.0959 Intercept
p              1  -1.17654 0.605012   -1.94   0.0568 Price
y              1  0.404115 0.117179    3.45   0.0011 Income
s              1  0.359204 0.085077    4.22   <.0001 Price of Substitutes


                        Test Results for Variable TEST_1

                     Num DF      Den DF    F Value    Pr > F

                          1         113       6.79    0.0104
```

Figure 19.10. TEST Statement Results

To test hypotheses involving parameters in different equations, use the STEST statement. Specify the parameters in the linear hypothesis as *model-label.regressor-name*. (If the MODEL statement does not have a label, you can use the dependent variable name as the label for the model, provided the dependent variable uniquely labels the model.)

For example, the following statements test the hypothesis that the income coefficient in the demand equation is .01 times the unit cost coefficient in the supply equation:

```
proc syslin data=in 3sls;
   endogenous  p;
   instruments y u s;
   demand: model q = p y s;
   supply: model q = p u;
   stest1: stest demand.y = .01 * supply.u;
run;
```

The STEST statement results are shown in Figure 19.11. The form and interpretation of the STEST statement results is like the TEST statement results. In this case, the *F*-test produces a *p*-value less than .0001, and strongly rejects the hypothesis tested. See the section "STEST Statement" for more information.

```
                              The SYSLIN Procedure
                      Three-Stage Least Squares Estimation

                     System Weighted MSE            0.5711
                     Degrees of freedom                113
                     System Weighted R-Square       0.9627


                         Model                   DEMAND
                         Dependent Variable           q
                         Label                 Quantity


                              Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  1.980261 1.169169    1.69   0.0959 Intercept
p              1  -1.17654 0.605012   -1.94   0.0568 Price
y              1  0.404115 0.117179    3.45   0.0011 Income
s              1  0.359204 0.085077    4.22   <.0001 Price of Substitutes


                         Model                   SUPPLY
                         Dependent Variable           q
                         Label                 Quantity


                              Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  -0.51878 0.490999   -1.06   0.2952 Intercept
p              1  1.333080 0.059271   22.49   <.0001 Price
u              1  -1.14623 0.243491   -4.71   <.0001 Unit Cost


                        Test Results for Variable STEST1

                     Num DF      Den DF    F Value    Pr > F

                          1         113      22.46    0.0001
```

Figure 19.11. STEST Statement Results

You can combine TEST and STEST statements with RESTRICT and SRESTRICT statements to perform hypothesis tests for restricted models. Of course, the validity of the TEST and STEST statement results will depend on the correctness of any restrictions you impose on the estimates.

Saving Residuals and Predicted Values

You can store predicted values and residuals from the estimated models in a SAS data set. Specify the OUT= option on the PROC SYSLIN statement and use the OUTPUT statement to specify names for new variables to contain the predicted and residual values.

For example, the following statements store the predicted quantity from the supply and demand equations in a data set PRED:

```
proc syslin data=in out=pred 3sls;
   endogenous  p;
   instruments y u s;
   demand: model q = p y s;
   output predicted=q_demand;
   supply: model q = p u;
   output predicted=q_supply;
run;
```

Plotting Residuals

You can plot the residuals against the regressors by specifying the PLOT option on the MODEL statement. For example, the following statements plot the 2SLS residuals for the demand model against price, income, price of substitutes, and the intercept.

```
proc syslin data=in 2sls;
   endogenous  p;
   instruments y u s;
   demand: model q = p y s / plot;
run;
```

The plot for price is shown in Figure 19.12. The other plots are not shown.

```
         --+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+--
  DEMAND |                                                                     |



                              The SYSLIN Procedure
                       Two-Stage Least Squares Estimation

         |                                                                     |
         |                                                                     |
    0.6 +                                                                     +
         |                                                                     |
Q        |                                                                     |
u        |                                                                     |
a   0.4 +                                                                     +
n        |                                               1 1                   |
t        |                                                                     |
i        |                                       111                           |
t   0.2 +                                  2      1 1                         +
y        |                    1    1  11              1 1                      |
         |                   1   1  1                  1                       |
         |                     2  1       1    11 1       11                   |
    0.0 +                           1    11       2                           +
         |              1  1   1            1    1       1                     |
         |                 1       1    1  1  1  1   1 1                       |
         |                  1 1    1        1 1                                |
   -0.2 +                            1         2       1                      +
         |                       1                                             |
         |                   1           1                                     |
         |                                  1                                  |
   -0.4 +                                                                     +
         |                                                                     |
         |                                                                     |
         |                                                                     |
   -0.6 +                                                                     +
         |                                                                     |
         |                                                                     |
         --+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+--
          2.0   2.2   2.4   2.6   2.8   3.0   3.2   3.4   3.6   3.8   4.0   4.2

                            Price                  p
```

Figure 19.12. PLOT Option Output for P

Syntax

The SYSLIN procedure uses the following statements:

PROC SYSLIN *options* **;**
BY *variables* **;**
ENDOGENOUS *variables* **;**
IDENTITY *identities* **;**
INSTRUMENTS *variables* **;**
MODEL *response = regressors / options* **;**
OUTPUT PREDICTED= *variable* **RESIDUAL=** *variable* **;**
RESTRICT *restrictions* **;**
SRESTRICT *restrictions* **;**
STEST *equations* **;**
TEST *equations* **;**
VAR *variables* **;**
WEIGHT *variable* **;**

Functional Summary

The SYSLIN procedure statements and options are summarized in the following table.

Description	Statement	Option
Data Set Options		
specify the input data set	PROC SYSLIN	DATA=
specify the output data set	PROC SYSLIN	OUT=
write parameter estimates to an output data set	PROC SYSLIN	OUTEST=
write covariances to the OUTEST= data set	PROC SYSLIN	OUTCOV OUTCOV3
write the SSCP matrix to an output data set	PROC SYSLIN	OUTSSCP=
Estimation Method Options		
specify full information maximum likelihood estimation	PROC SYSLIN	FIML
specify iterative SUR estimation	PROC SYSLIN	ITSUR
specify iterative 3SLS estimation	PROC SYSLIN	IT3SLS
specify K-class estimation	PROC SYSLIN	K=
specify limited information maximum likelihood estimation	PROC SYSLIN	LIML
specify minimum expected loss estimation	PROC SYSLIN	MELO
specify ordinary least squares estimation	PROC SYSLIN	OLS
specify seemingly unrelated estimation	PROC SYSLIN	SUR
specify two-stage least-squares estimation	PROC SYSLIN	2SLS

Description	Statement	Option
specify three-stage least-squares estimation	PROC SYSLIN	3SLS
specify Fuller's modification to LIML	PROC SYSLIN	ALPHA=
specify convergence criterion	PROC SYSLIN	CONVERGE=
specify maximum number of iterations	PROC SYSLIN	MAXIT=
use diagonal of **S** instead of **S**	PROC SYSLIN	SDIAG
exclude RESTRICT statements in final stage	PROC SYSLIN	NOINCLUDE
specify criterion for testing for singularity	PROC SYSLIN	SINGULAR=
specify denominator for variance estimates	PROC SYSLIN	VARDEF=
Printing Control Options		
print first-stage regression statistics	PROC SYSLIN	FIRST
print estimates and SSE at each iteration	PROC SYSLIN	ITPRINT
print the restricted reduced form estimates	PROC SYSLIN	REDUCED
print descriptive statistics	PROC SYSLIN	SIMPLE
print uncorrected SSCP matrix	PROC SYSLIN	USSCP
print correlations of the parameter estimates	MODEL	CORRB
print covariances of the parameter estimates	MODEL	COVB
print Durbin-Watson statistics	MODEL	DW
print Basmann's test	MODEL	OVERID
plot residual values against regressors	MODEL	PLOT
print standardized parameter estimates	MODEL	STB
print unrestricted parameter estimates	MODEL	UNREST
print the model crossproducts matrix	MODEL	XPX
print the inverse of the crossproducts matrix	MODEL	I
suppress printed output	MODEL	NOPRINT
suppress all printed output	PROC SYSLIN	NOPRINT
Model Specification		
specify structural equations	MODEL	
suppress the intercept parameter	MODEL	NOINT
specify linear relationship among variables	IDENTITY	
perform weighted regression	WEIGHT	
Tests and Restrictions on Parameters		
place restrictions on parameter estimates	RESTRICT	
place restrictions on parameter estimates	SRESTRICT	
test linear hypothesis	STEST	
test linear hypothesis	TEST	
Other Statements		
specify BY-group processing	BY	

Description	Statement	Option
specify the endogenous variables	ENDOGENOUS	
specify instrumental variables	INSTRUMENTS	
write predicted and residual values to a data set	OUTPUT	
name variable for predicted values	OUTPUT	PREDICTED=
name variable for residual values	OUTPUT	RESIDUAL=
include additional variables in $X'X$ matrix	VAR	

PROC SYSLIN Statement

PROC SYSLIN *options;*

The following options can be used with the PROC SYSLIN statement.

Data Set Options

DATA= *SAS-data-set*
specifies the input data set. If the DATA= option is omitted, the most recently created SAS data set is used. In addition to ordinary SAS data sets, PROC SYSLIN can analyze data sets of TYPE=CORR, TYPE=COV, TYPE=UCORR, TYPE=UCOV, and TYPE=SSCP. See "Special TYPE= Input Data Set" in the "Input Data Set" section later in this chapter for more information.

OUT= *SAS-data-set*
specifies an output SAS data set for residuals and predicted values. The OUT= option is used in conjunction with the OUTPUT statement. See the section "OUT= Data Set" later in this chapter for more details.

OUTEST= *SAS-data-set*
writes the parameter estimates to an output data set. See the section "OUTEST= Data Set" later in this chapter for details.

OUTCOV
COVOUT
writes the covariance matrix of the parameter estimates to the OUTEST= data set in addition to the parameter estimates.

OUTCOV3
COV3OUT
writes covariance matrices for each model in a system to the OUTEST= data set when the 3SLS, SUR, or FIML option is used.

OUTSSCP= *SAS-data-set*
writes the sum-of-squares-and-crossproducts matrix to an output data set. See the section "OUTSSCP= Data Set" later in this chapter for details.

Estimation Method Options

2SLS
specifies the two-stage least-squares estimation method.

3SLS
specifies the three-stage least-squares estimation method.

FIML
specifies the full information maximum likelihood estimation method.

ITSUR
specifies the iterative seemingly unrelated estimation method.

IT3SLS
specifies the iterative three-stage least-squares estimation method.

K= *value*
specifies the K-class estimation method.

LIML
specifies the limited information maximum likelihood estimation method.

MELO
specifies the minimum expected loss estimation method.

OLS
specifies the ordinary least squares estimation method. This is the default.

SUR
specifies the seemingly unrelated estimation method.

Printing and Control Options

ALL
specifies the CORRB, COVB, DW, I, OVERID, PLOT, STB, and XPX options for every MODEL statement.

ALPHA= *value*
specifies Fuller's modification to the LIML estimation method. See "Fuller's Modification to LIML *K* Value" later in this chapter for details.

CONVERGE= *value*
specifies the convergence criterion for the iterative estimation methods IT3SLS, ITSUR, and FIML. The default is CONVERGE=.0001.

FIRST
prints first-stage regression statistics for the endogenous variables regressed on the instruments. This output includes sums of squares, estimates, variances, and standard deviations.

ITPRINT
prints parameter estimates, system-weighted residual sum of squares, and R^2 at each iteration for the IT3SLS and ITSUR estimation methods. For the FIML method, the ITPRINT option prints parameter estimates, negative of log likelihood function, and norm of gradient vector at each iteration.

MAXITER= *n*

specifies the maximum number of iterations allowed for the IT3SLS, ITSUR, and FIML estimation methods. The MAXITER= option can be abbreviated as MAXIT=. The default is MAXITER=30.

NOINCLUDE

excludes the RESTRICT statements from the final stage for the 3SLS, IT3SLS, SUR, ITSUR estimation methods.

NOPRINT

suppresses all printed output. Specifying NOPRINT in the PROC SYSLIN statement is equivalent to specifying NOPRINT in every MODEL statement.

REDUCED

prints the reduced form estimates. If the REDUCED option is specified, you should specify any IDENTITY statements needed to make the system square. See "Reduced Form Estimates" in the section "Computational Details" later in this chapter for more information.

SDIAG

uses the diagonal of **S** instead of **S** to do the estimation, where **S** is the covariance matrix of equation errors. See "Uncorrelated Errors Across Equations" in the section "Computational Details" later in this chapter for more information.

SIMPLE

prints descriptive statistics for the dependent variables. The statistics printed include the sum, mean, uncorrected sum of squares, variance, and standard deviation.

SINGULAR= *value*

specifies a criterion for testing singularity of the crossproducts matrix. This is a tuning parameter used to make PROC SYSLIN more or less sensitive to singularities. The value must be between 0 and 1. The default is SINGULAR=1E-8.

USSCP

prints the uncorrected sum-of-squares-and-crossproducts matrix.

USSCP2

prints the uncorrected sum-of-squares-and-crossproducts matrix for all variables used in the analysis, including predicted values of variables generated by the procedure.

VARDEF= DF | N | WEIGHT | WGT

specifies the denominator to use in calculating cross-equation error covariances and parameter standard errors and covariances. The default is VARDEF=DF, which corrects for model degrees of freedom. VARDEF=N specifies no degrees-of-freedom correction. VARDEF=WEIGHT specifies the sum of the observation weights. VARDEF=WGT specifies the sum of the observation weights minus the model degrees of freedom. See "Computation of Standard Errors" in the section "Computational Details" later in this chapter for more information.

BY Statement

BY *variables ;*

A BY statement can be used with PROC SYSLIN to obtain separate analyses on observations in groups defined by the BY variables.

ENDOGENOUS Statement

ENDOGENOUS *variables ;*

The ENDOGENOUS statement declares the jointly dependent variables that are projected in the first-stage regression through the instrument variables. The ENDOGENOUS statement is not needed for the SUR, ITSUR, or OLS estimation methods. The default ENDOGENOUS list consists of all the dependent variables in the MODEL and IDENTITY statements that do not appear in the INSTRUMENTS statement.

IDENTITY Statement

IDENTITY *equation ;*

The IDENTITY statement specifies linear relationships among variables to write to the OUTEST= data set. It provides extra information in the OUTEST= data set but does not create or compute variables. The OUTEST= data set can be processed by the SIMLIN procedure in a later step.

The IDENTITY statement is also used to compute reducedform coefficients when the REDUCED option in the PROC SYSLIN statement is specified. See "Reduced Form Estimates" in the section "Computational Details" later in this chapter for more information.

The *equation* given by the IDENTITY statement has the same form as equations in the MODEL statement. A label can be specified for an IDENTITY statement as follows:

*label***: IDENTITY** . . . **;**

INSTRUMENTS Statement

INSTRUMENTS *variables ;*

The INSTRUMENTS statement declares the variables used in obtaining first-stage predicted values. All the instruments specified are used in each first-stage regression. The INSTRUMENTS statement is required for the 2SLS, 3SLS, IT3SLS, LIML, MELO, and K-class estimation methods. The INSTRUMENTS statement is not needed for the SUR, ITSUR, OLS, or FIML estimation methods.

MODEL Statement

MODEL *response = regressors / options ;*

The MODEL statement regresses the response variable on the left side of the equal sign against the regressors listed on the right side.

Models can be given labels. Model labels are used in the printed output to identify the results for different models. Model labels are also used in SRESTRICT and STEST statements to refer to parameters in different models. If no label is specified, the response variable name is used as the label for the model. The model label is specified as follows:

label: **MODEL** . . . **;**

The following options can be used in the MODEL statement after a slash (/).

ALL
specifies the CORRB, COVB, DW, I, OVERID, PLOT, STB, and XPX options.

ALPHA= *value*
specifies the α parameter for Fuller's modification to the LIML estimation method. See "Fuller's Modification to LIML" in the section "Computational Details" later in this chapter for more information.

CORRB
prints the matrix of estimated correlations between the parameter estimates.

COVB
prints the matrix of estimated covariances between the parameter estimates.

DW
prints Durbin-Watson statistics and autocorrelation coefficients for the residuals. If there are missing values, d' is calculated according to Savin and White (1978). Use the DW option only if the data set to be analyzed is an ordinary SAS data set with time series observations sorted in time order. The Durbin-Watson test is not valid for models with lagged dependent regressors.

I
prints the inverse of the crossproducts matrix for the model, $(\mathbf{X}'\mathbf{X})^{-1}$. If restrictions are specified, the crossproducts matrix printed is adjusted for the restrictions. See the section "Computational Details" for more information.

K= *value*
specifies K-class estimation.

NOINT
suppresses the intercept parameter from the model.

NOPRINT
suppresses the normal printed output.

OVERID
prints Basmann's (1960) test for over identifying restrictions. See "Over Identification Restrictions" in the section "Computational Details" later in this chapter for more information.

PLOT
plots residual values against regressors. A plot of the residuals for each regressor is printed.

STB
prints standardized parameter estimates. Sometimes known as a standard partial regression coefficient, a standardized parameter estimate is a parameter estimate multiplied by the standard deviation of the associated regressor and divided by the standard deviation of the response variable.

UNREST
prints parameter estimates computed before restrictions are applied. The UNREST option is valid only if a RESTRICT statement is specified.

XPX
prints the model crossproducts matrix, $X'X$. See the section "Computational Details" for more information.

OUTPUT Statement

OUTPUT PREDICTED=*variable* **RESIDUAL**=*variable* **;**

The OUTPUT statement writes predicted values and residuals from the preceding model to the data set specified by the OUT= option on the PROC SYSLIN statement. An OUTPUT statement must come after the MODEL statement to which it applies. The OUT= option must be specified in the PROC SYSLIN statement.

The following options can be specified in the OUTPUT statement:

PREDICTED= *variable*
names a new variable to contain the predicted values for the response variable. The PREDICTED= option can be abbreviated as PREDICT=, PRED=, or P=.

RESIDUAL= *variable*
names a new variable to contain the residual values for the response variable. The RESIDUAL= option can be abbreviated as RESID= or R=.

For example, the following statements create an output data set named B. In addition to the variables in the input data set, the data set B contains the variable YHAT, with values that are predicted values of the response variable Y, and YRESID, with values that are the residual values of Y.

```
proc syslin data=a out=b;
   model y = x1 x2;
   output p=yhat r=yresid;
run;
```

For example, the following statements create an output data set named PRED. In addition to the variables in the input data set, the data set PRED contains the variables Q_DEMAND and Q_SUPPLY, with values that are predicted values of the response variable Q for the demand and supply equations respectively, and R_DEMAND and R_SUPPLY, with values that are the residual values of the demand and supply equations.

```
proc syslin data=in out=pred;
   demand: model q = p y s;
   output p=q_demand r=r_demand;
   supply: model q = p u;
   output p=q_supply r=r_supply;
run;
```

See the section "OUT= Data Set" later in this chapter for more details.

RESTRICT Statement

RESTRICT *equation , ... , equation ;*

The RESTRICT statement places restrictions on the parameter estimates for the preceding MODEL statement. Any number of restrict statements can follow a MODEL statement. Each restriction is written as a linear equation. If more than one restriction is specified in a single RESTRICT statement, the restrictions are separated by commas.

Parameters are referred to by the name of the corresponding regressor variable. Each name used in the equation must be a regressor in the preceding MODEL statement. The keyword INTERCEPT is used to refer to the intercept parameter in the model.

RESTRICT statements can be given labels. The labels are used in the printed output to distinguish results for different restrictions. Labels are specified as follows:

label **: RESTRICT** . . . **;**

The following is an example of the use of the RESTRICT statement, in which the coefficients of the regressors X1 and X2 are required to sum to 1.

```
proc syslin data=a;
   model y = x1 x2;
   restrict x1 + x2 = 1;
run;
```

Variable names can be multiplied by constants. When no equal sign appears, the linear combination is set equal to 0. Note that the parameters associated with the variables are restricted, not the variables themselves. Here are some examples of valid RESTRICT statements:

```
restrict x1 + x2 = 1;
restrict x1 + x2 - 1;
restrict 2 * x1 = x2 + x3 , intercept + x4 = 0;
restrict x1 = x2 = x3 = 1;
restrict 2 * x1 - x2;
```

Restricted parameter estimates are computed by introducing a Lagrangian parameter λ for each restriction (Pringle and Raynor 1971). The estimates of these Lagrangian parameters are printed in the parameter estimates table. If a restriction cannot be applied, its parameter value and degrees of freedom are listed as 0.

The Lagrangian parameter, λ, measures the sensitivity of the SSE to the restriction. If the restriction is changed by a small amount ϵ, the SSE is changed by $2\lambda\epsilon$.

The *t*-ratio tests the significance of the restrictions. If λ is zero, the restricted estimates are the same as the unrestricted.

Any number of restrictions can be specified on a RESTRICT statement, and any number of RESTRICT statements can be used. The estimates are computed subject to all restrictions specified. However, restrictions should be consistent and not redundant.

Note: The RESTRICT statement is not supported for the FIML estimation method.

SRESTRICT Statement

SRESTRICT *equation , ... , equation ;*

The SRESTRICT statement imposes linear restrictions involving parameters in two or more MODEL statements. The SRESTRICT statement is like the RESTRICT statement but is used to impose restrictions across equations, whereas the RESTRICT statement only applies to parameters in the immediately preceding MODEL statement.

Each restriction is written as a linear equation. Parameters are referred to as *label.variable,* where *label* is the model label and *variable* is the name of the regressor to which the parameter is attached. (If the MODEL statement does not have a label, you can use the dependent variable name as the label for the model, provided the dependent variable uniquely labels the model.) Each variable name used must be a regressor in the indicated MODEL statement. The keyword INTERCEPT is used to refer to intercept parameters.

SRESTRICT statements can be given labels. The labels are used in the printed output to distinguish results for different restrictions. Labels are specified as follows:

label **: SRESTRICT** . . . **;**

The following is an example of the use of the SRESTRICT statement, in which the coefficient for the regressor X2 is constrained to be the same in both models.

```
proc syslin data=a 3sls;
   endogenous y1 y2;
   instruments x1 x2;
   model y1 = y2 x1 x2;
   model y2 = y1 x2;
   srestrict y1.x2 = y2.x2;
run;
```

When no equal sign is used, the linear combination is set equal to 0. Thus the restriction in the preceding example can also be specified as

```
   srestrict y1.x2 - y2.x2;
```

Any number of restrictions can be specified on an SRESTRICT statement, and any number of SRESTRICT statements can be used. The estimates are computed subject to all restrictions specified. However, restrictions should be consistent and not redundant.

The results of the SRESTRICT statements are printed after the parameter estimates for all the models in the system. The format of the SRESTRICT statement output is the same as the parameter estimates table. In this output the "Parameter Estimate" is the Lagrangian parameter, λ, used to impose the restriction.

The Lagrangian parameter, λ, measures the sensitivity of the system sum of square errors to the restriction. The system SSE is the system MSE shown in the printed output multiplied by the degrees of freedom. If the restriction is changed by a small amount ϵ, the system SSE is changed by $2\lambda\epsilon$.

The *t*-ratio tests the significance of the restriction. If λ is zero, the restricted estimates are the same as the unrestricted estimates.

The model degrees of freedom are not adjusted for the cross-model restrictions imposed by SRESTRICT statements.

Note: The SRESTRICT statement is not supported for the FIML estimation method.

STEST Statement

STEST *equation , ... , equation / options ;*

The STEST statement performs an *F*-test for the joint hypotheses specified in the statement.

The hypothesis is represented in matrix notation as

$$\mathbf{L}\beta = \mathbf{c}$$

and the *F*-test is computed as

$$\frac{(\mathbf{L}b - \mathbf{c})'(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{L}')^{-1}(\mathbf{L}b - \mathbf{c})}{m\hat{\sigma}^2}$$

where b is the estimate of β, m is the number of restrictions, and $\hat{\sigma}^2$ is the system weighted mean square error. See the section "Computational Details" for information on the matrix $\mathbf{X}'\mathbf{X}$.

Each hypothesis to be tested is written as a linear equation. Parameters are referred to as *label.variable*, where *label* is the model label and *variable* is the name of the regressor to which the parameter is attached. (If the MODEL statement does not have a label, you can use the dependent variable name as the label for the model, provided the dependent variable uniquely labels the model.) Each variable name used must be a regressor in the indicated MODEL statement. The keyword INTERCEPT is used to refer to intercept parameters.

STEST statements can be given labels. The label is used in the printed output to distinguish different tests. Any number of STEST statements can be specified. Labels are specified as follows:

*label***: STEST** . . . **;**

The following is an example of the STEST statement:

```
proc syslin data=a 3sls;
   endogenous y1 y2;
   instruments x1 x2;
   model y1 = y2 x1 x2;
   model y2 = y1 x2;
   stest y1.x2 = y2.x2;
run;
```

The test performed is exact only for ordinary least squares, given the OLS assumptions of the linear model. For other estimation methods, the *F*-test is based on large sample theory and is only approximate in finite samples.

If RESTRICT or SRESTRICT statements are used, the tests computed by the STEST statement are conditional on the restrictions specified. The validity of the tests may be compromised if incorrect restrictions are imposed on the estimates.

The following are examples of STEST statements:

```
stest a.x1 + b.x2 = l;
stest 2 * b.x2 = c.x3 + c.x4 ,
      a.intercept + b.x2 = 0;
stest a.x1 = c.x2 = b.x3 = 1;
stest 2 * a.x1 - b.x2 = 0;
```

The PRINT option can be specified in the STEST statement after a slash (/):

PRINT
prints intermediate calculations for the hypothesis tests.

Note: The STEST statement is not supported for the FIML estimation method.

TEST Statement

TEST *equation , ... , equation / options ;*

The TEST statement performs *F*-tests of linear hypotheses about the parameters in the preceding MODEL statement. Each equation specifies a linear hypothesis to be tested. If more than one equation is specified, the equations are separated by commas.

Variable names must correspond to regressors in the preceding MODEL statement, and each name represents the coefficient of the corresponding regressor. The keyword INTERCEPT is used to refer to the model intercept.

TEST statements can be given labels. The label is used in the printed output to distinguish different tests. Any number of TEST statements can be specified. Labels are specified as follows:

label: **TEST** . . . **;**

The following is an example of the use of TEST statement, which tests the hypothesis that the coefficients of X1 and X2 are the same:

```
proc syslin data=a;
   model y = x1 x2;
   test x1 = x2;
run;
```

The following statements perform *F*-tests for the hypothesis that the coefficients of X1 and X2 are equal, and that the sum of the X1 and X2 coefficients is twice the intercept, and for the joint hypothesis.

```
proc syslin data=a;
   model y = x1 x2;
   x1eqx2:  test x1 = x2;
   sumeq2i: test x1 + x2 = 2 * intercept;
   joint:   test x1 = x2, x1 + x2 = 2 * intercept;
run;
```

The following are additional examples of TEST statements:

```
test x1 + x2 = 1;
test x1 = x2 = x3 = 1;
test 2 * x1 = x2 + x3, intercept + x4 = 0;
test 2 * x1 - x2;
```

The TEST statement performs an *F*-test for the joint hypotheses specified. The hypothesis is represented in matrix notation as follows:

$$\mathbf{L}\beta = \mathbf{c}$$

The *F* test is computed as

$$\frac{(\mathbf{L}b - \mathbf{c})'(\mathbf{L}(\mathbf{X}'\mathbf{X})^{-}\mathbf{L}')^{-1}(\mathbf{L}b - \mathbf{c})}{m\hat{\sigma}^2}$$

where b is the estimate of β, m is the number of restrictions, and $\hat{\sigma}^2$ is the model mean square error. See the section "Computational Details" for information on the matrix $\mathbf{X}'\mathbf{X}$.

The test performed is exact only for ordinary least squares, given the OLS assumptions of the linear model. For other estimation methods, the *F*-test is based on large sample theory and is only approximate in finite samples.

If RESTRICT or SRESTRICT statements are used, the tests computed by the TEST statement are conditional on the restrictions specified. The validity of the tests may be compromised if incorrect restrictions are imposed on the estimates.

The PRINT option can be specified in the TEST statement after a slash (/):

PRINT

prints intermediate calculations for the hypothesis tests.

Note: The TEST statement is not supported for the FIML estimation method.

VAR Statement

VAR *variables ;*

The VAR statement is used to include variables in the crossproducts matrix that are not specified in any MODEL statement. This statement is rarely used with PROC SYSLIN and is used only with the OUTSSCP= option in the PROC SYSLIN statement.

WEIGHT Statement

WEIGHT *variable ;*

The WEIGHT statement is used to perform weighted regression. The WEIGHT statement names a variable in the input data set whose values are relative weights for a weighted least-squares fit. If the weight value is proportional to the reciprocal of the variance for each observation, the weighted estimates are the best linear unbiased estimates (BLUE).

Details

Input Data Set

PROC SYSLIN does not compute new values for regressors. For example, if you need a lagged variable, you must create it with a DATA step. No values are computed by IDENTITY statements; all values must be in the input data set.

Special TYPE= Input Data Set

The input data set for most applications of the SYSLIN procedure contains standard rectangular data. However, PROC SYSLIN can also process input data in the form of a crossproducts, covariance, or correlation matrix. Data sets containing such matrices are identified by values of the TYPE= data set option.

These special kinds of input data sets can be used to save computer time. It takes nk^2 operations, where n is the number of observations and k is the number of variables, to calculate cross products; the regressions are of the order k^3. When n is in the thousands and k is much smaller, you can save most of the computer time in later runs of PROC SYSLIN by re-using the SSCP matrix rather than recomputing it.

The SYSLIN procedure can process TYPE= CORR, COV, UCORR, UCOV, or SSCP data sets. TYPE=CORR and TYPE=COV data sets, usually created by the CORR procedure, contain means and standard deviations, and correlations or covariances. TYPE=SSCP data sets, usually created in previous runs of PROC SYSLIN, contain sums of squares and cross products. Refer to *SAS/STAT User's Guide* for more information on special SAS data sets.

When special SAS data sets are read, you must specify the TYPE= data set option. PROC CORR and PROC SYSLIN automatically set the type for output data sets; however, if you create the data set by some other means, you must specify its type with the TYPE= data set option.

When the special data sets are used, the DW (Durbin-Watson test) and PLOT options in the MODEL statement cannot be performed, and the OUTPUT statements are not valid.

Estimation Methods

A brief description of the methods used by the SYSLIN procedure follows. For more information on these methods, see the references at the end of this chapter.

There are two fundamental methods of estimation for simultaneous equations: least squares and maximum likelihood. There are two approaches within each of these categories: single equation methods and system estimation. 2SLS, 3SLS, and IT3SLS use the least-squares method; LIML and FIML use the maximum likelihood method. 2SLS and LIML are single equation methods, which means that over identifying restrictions in other equations are not taken into account in estimating parameters in a particular equation. (See "Over Identification Restrictions" in the section "Computational Details" later in this chapter for more information.) As a result, 2SLS and LIML estimates are not asymptotically efficient. The system methods are 3SLS, IT3SLS, and FIML. These methods use information concerning the endogenous variables in the system and take into account error covariances across equations and hence are asymptotically efficient in the absence of specification error.

K-class estimation is a class of estimation methods that include the 2SLS, OLS, LIML, and MELO methods as special cases. A *K*-value less than 1 is recommended but not required.

MELO is a Bayesian K-class estimator. It yields estimates that can be expressed as a matrix weighted average of the OLS and 2SLS estimates.

The SUR and ITSUR methods use information about contemporaneous correlation among error terms across equations in an attempt to improve the efficiency of parameter estimates.

Instrumental Variables and K-Class Estimation Methods

Instrumental variable methods involve substituting a predicted variable for the endogenous variable Y when it appears as a regressor. The predicted variables are linear functions of the instrumental variables and the endogenous variable.

The 2SLS method substitutes $\hat{Y}$ for Y, which results in consistent estimates. In 2SLS, the instrumental variables are used as regressors to obtain the projected value $\hat{Y}$, which is then substituted for Y. Normally, the predetermined variables of the system are used as the instruments. It is possible to use variables other than predetermined variables from your system of equations as instruments; however, the estimation may not be as efficient. For consistent estimates, the instruments must be uncorrelated with the residual and correlated with the endogenous variable.

K-class estimators are instrumental variable estimators where the first-stage predicted values take a special form: $Y^* = (1-k)Y + k\hat{Y}$ for a specified value *k*. The probability limit of *k* must equal 1 for consistent parameter estimates.

The LIML method results in consistent estimates that are exactly equal to 2SLS estimates when an equation is exactly identified. LIML can be viewed as least-variance ratio estimators or as maximum likelihood estimators. LIML involves minimizing the ratio $\lambda = (rvar\_eq)/(rvar\_sys)$, where $rvar\_eq$ is the residual variance associated with regressing the weighted endogenous variables on all predetermined variables appearing in that equation, and $rvar\_sys$ is the residual variance associated with regressing weighted endogenous variables on all predetermined variables in the system. The K-class interpretation of LIML is that $k = \lambda$. Unlike OLS and 2SLS, where *k* is 0 and 1, respectively, *k* is stochastic in the LIML method.

The MELO method computes the minimum expected loss estimator. The MELO method computes estimates that "minimize the posterior expectation of generalized quadratic loss functions for structural coefficients of linear structural models" (Judge et al. 1985, 635). Other frequently used K-class estimators may not have finite moments under some commonly encountered circumstances and hence there can be infinite risk relative to quadratic and other loss functions. MELO estimators have finite second moments and hence finite risk.

One way of comparing K-class estimators is to note that when k=1, the correlation between regressor and the residual is completely corrected for. In all other cases, it is only partially corrected for.

SUR and 3SLS Estimation Methods

SUR may improve the efficiency of parameter estimates when there is contemporaneous correlation of errors across equations. In practice, the contemporaneous correlation matrix is estimated using OLS residuals. Under two sets of circumstances, SUR parameter estimates are the same as those produced by OLS: when there is no contemporaneous correlation of errors across equations (the estimate of contemporaneous correlation matrix is diagonal); and when the independent variables are the same across equations.

Theoretically, SUR parameter estimates will always be at least as efficient as OLS in large samples, provided that your equations are correctly specified. However, in small samples the need to estimate the covariance matrix from the OLS residuals increases the sampling variability of the SUR estimates, and this effect can cause SUR to be less efficient than OLS. If the sample size is small and the across-equation correlations are small, then OLS should be preferred to SUR. The consequences of specification error are also more serious with SUR than with OLS.

The 3SLS method combines the ideas of the 2SLS and SUR methods. Like 2SLS, the 3SLS method uses $\hat{Y}$ instead of Y for endogenous regressors, which results in consistent estimates. Like SUR, the 3SLS method takes the cross-equation error correlations into account to improve large sample efficiency. For 3SLS, the 2SLS residuals are used to estimate the cross-equation error covariance matrix.

The SUR and 3SLS methods can be iterated by recomputing the estimate of the cross-equation covariance matrix from the SUR or 3SLS residuals and then computing new SUR or 3SLS estimates based on this updated covariance matrix estimate. Continuing this iteration until convergence produces ITSUR or IT3SLS estimates.

FIML Estimation Method

The FIML estimator is a system generalization of the LIML estimator. The FIML method involves minimizing the determinant of the covariance matrix associated with residuals of the reduced form of the equation system. From a maximum likelihood standpoint, the LIML method involves assuming that the errors are normally distributed and then maximizing the likelihood function subject to restrictions on a particular equation. FIML is similar, except that the likelihood function is maximized subject to restrictions on all of the parameters in the model, not just those in the equation being estimated. The FIML method is implemented as an instrumental variable method (Hausman 1975).

Note: the RESTRICT, SRESTRICT, TEST, and STEST statements are not supported when the FIML method is used.

Choosing a Method for Simultaneous Equations

A number of factors should be taken into account in choosing an estimation method. Although system methods are asymptotically most efficient in the absence of specification error, system methods are more sensitive to specification error than single equation methods.

In practice, models are never perfectly specified. It is a matter of judgment whether the misspecification is serious enough to warrant avoidance of system methods.

Another factor to consider is sample size. With small samples, 2SLS may be preferred to 3SLS. In general, it is difficult to say much about the small sample properties of K-class estimators because this depends on the regressors used.

LIML and FIML are invariant to the normalization rule imposed but are computationally more expensive than 2SLS or 3SLS.

If the reason for contemporaneous correlation among errors across equations is a common omitted variable, it is not necessarily best to apply SUR. SUR parameter estimates are more sensitive to specification error than OLS. OLS may produce better parameter estimates under these circumstances. SUR estimates are also affected by the sampling variation of the error covariance matrix. There is some evidence from Monte Carlo studies that SUR is less efficient than OLS in small samples.

ANOVA Table for Instrumental Variables Methods

In the instrumental variables methods (2SLS, LIML, K-class, MELO), first-stage predicted values are substituted for the endogenous regressors. As a result, the regression sum of squares (RSS) and the error sum of squares (ESS) do not sum to the total corrected sum of squares for the dependent variable (TSS). The "Analysis of Variance" table printed for the second-stage results serves to display these sums of squares and the mean squares used for the *F*-test, but this table is not a variance decomposition in the usual analysis of variance sense.

The *F*-test shown in the instrumental variables case is a valid test of the no-regression hypothesis that the true coefficients of all regressors are 0. However, because of the first-stage projection of the regression mean square, this is a Wald-type test statistic, which is asymptotically *F* but not exactly F-distributed in finite samples. Thus, for small samples the *F*-test is only approximate when instrumental variables are used.

The R^2 Statistics

As explained in the section "ANOVA Table for Instrumental Variables Methods," when instrumental variables are used, the regression sum of squares (RSS) and the error sum of squares (ESS) do not sum to the total corrected sum of squares. In this case, there are several ways that the R^2 statistic can be defined.

The definition of R^2 used by the SYSLIN procedure is

$$\mathrm{R}^2 = \frac{\mathrm{RSS}}{\mathrm{RSS} + \mathrm{ESS}}$$

This definition is consistent with the *F*-test of the null hypothesis that the true coefficients of all regressors are zero. However, this R^2 may not be a good measure of the goodness of fit of the model.

System Weighted R^2 and System Weighted Mean Square Error

The system weighted R^2, printed for the 3SLS, IT3SLS, SUR, ITSUR, and FIML methods, is computed as follows.

$$R^2 = \mathbf{Y'WR(X'X)^{-1}R'WY/Y'WY}$$

In this equation the matrix **X'X** is $\mathbf{R'WR}$, and **W** is the projection matrix of the instruments:

$$\mathbf{W} = \mathbf{S}^{-1} \otimes \mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'}$$

The matrix **Z** is the instrument set, **R** is the the regressor set, and **S** is the estimated cross-model covariance matrix.

The system weighted MSE, printed for the 3SLS, IT3SLS, SUR, ITSUR, and FIML methods, is computed as follows:

$$MSE = \frac{1}{tdf}(\mathbf{Y'WY} - \mathbf{Y'WR}(\mathbf{X'X})^{-1}\mathbf{R'WY})$$

In this equation *tdf* is the sum of the error degrees of freedom for the equations in the system.

Computational Details

This section discusses various computational details.

Computation of Model Crossproduct Matrix

Model crossproduct matrix $\mathbf{X'X}$ is formed from projected values. For K-class estimation,

$$\mathbf{X} = (1-k)\mathbf{R} + k\mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'R}$$

where **Z** is the instrument set and **R** is the the regressor set. Note that k=1 for the 2SLS method and k=0 for the OLS method.

In the 3SLS, IT3SLS, SUR, and ITSUR methods, $\mathbf{X'X}$ is formed as

$$\mathbf{X'X} = \mathbf{R'}(\mathbf{S}^{-1} \otimes \mathbf{Z}(\mathbf{Z'Z})^{-1}\mathbf{Z'})\mathbf{R}$$

where **Z** and **R** are as defined previously and **S** is an estimate of the cross-equation covariance matrix. For SUR and ITSUR, **Z** is the identity matrix.

Computation of Standard Errors

The VARDEF= option in the PROC SYSLIN statement controls the denominator used in calculating the cross-equation covariance estimates and the parameter standard errors and covariances. The values of the VARDEF= option and the resulting denominator are as follows:

N uses the number of nonmissing observations.

DF uses the number of nonmissing observations less the degrees of freedom in the model.

WEIGHT	uses the sum of the observation weights given by the WEIGHTS statement.
WDF	uses the sum of the observation weights given by the WEIGHTS statement less the degrees of freedom in the model.

The VARDEF= option does not affect the model mean square error, root mean square error, or R^2 statistics. These statistics are always based on the error degrees of freedom, regardless of the VARDEF= option. The VARDEF= option also does not affect the dependent variable coefficient of variation (C.V.).

Reduced Form Estimates

The REDUCED option on the PROC SYSLIN statement computes estimates of the reduced form coefficients. The REDUCED option requires that the equation system be square. If there are fewer models than endogenous variables, IDENTITY statements can be used to complete the equation system.

The reduced form coefficients are computed as follows. Represent the equation system, with all endogenous variables moved to the left-hand side of the equations and identities, as

$$\mathbf{BY} = \mathbf{\Gamma X}$$

Here **B** is the estimated coefficient matrix for the endogenous variables **Y**, and $\mathbf{\Gamma}$ is the estimated coefficient matrix for the exogenous (or predetermined) variables **X**.

The system can be solved for **Y** as follows, provided **B** is square and nonsingular:

$$\mathbf{Y} = \mathbf{B}^{-1}\mathbf{\Gamma X}$$

The reduced form coefficients are the matrix $\mathbf{B}^{-1}\mathbf{\Gamma}$.

Uncorrelated Errors Across Equations

The SDIAG option in the PROC SYSLIN statement computes estimates assuming uncorrelated errors across equations. As a result, when the SDIAG option is used, the 3SLS estimates are identical to 2SLS estimates, and the SUR estimates are the same as the OLS estimates.

Over Identification Restrictions

The OVERID option in the MODEL statement can be used to test for over identifying restrictions on parameters of each equation. The null hypothesis is that the predetermined variables not appearing in any equation have zero coefficients. The alternative hypothesis is that at least one of the assumed zero coefficients is nonzero. The test is approximate and rejects the null hypothesis too frequently for small sample sizes.

The formula for the test is given as follows. Let $y_i = \beta_i \mathbf{Y}_i + \gamma_i \mathbf{Z}_i + e_i$ be the *i*th equation. $\mathbf{Y}_i$ are the endogenous variables that appear as regressors in the *i*th equation, and $\mathbf{Z}_i$ are the instrumental variables that appear as regressors in the *i*th equation. Let N_i be the number of variables in $\mathbf{Y}_i$ and $\mathbf{Z}_i$.

Let $v_i = y_i - \mathbf{Y}_i \hat{\beta}_i$. Let $\mathbf{Z}$ represent all instrumental variables, *T* be the total number of observations, and *K* be the total number of instrumental variables. Define $\hat{l}$ as follows:

$$\hat{l} = \frac{v'_i(\mathbf{I} - \mathbf{Z}_i(\mathbf{Z}'_i\mathbf{Z}_i)^{-1}\mathbf{Z}'_i)v_i}{v'_i(\mathbf{I} - \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}')v_i}$$

Then the test statistic

$$\frac{T-K}{K-N_i}(\hat{l}-1)$$

is distributed approximately as an *F* with $K - N_i$ and $T - K$ degrees of freedom. Refer to Basmann (1960) for more information.

Fuller's Modification to LIML

The ALPHA= option in the PROC SYSLIN and MODEL statements parameterizes Fuller's modification to LIML. This modification is $k = \gamma - (\alpha/(n-g))$, where α is the value of the ALPHA= option, γ is the LIML *k* value, *n* is the number of observations, and *g* is the number of predetermined variables. Fuller's modification is not used unless the ALPHA= option is specified. Refer to Fuller (1977) for more information.

Missing Values

Observations having a missing value for any variable in the analysis are excluded from the computations.

OUT= Data Set

The output SAS data set produced by the OUT= option in the PROC SYSLIN statement contains all the variables in the input data set and the variables containing predicted values and residuals specified by OUTPUT statements.

The residuals are computed as actual values minus predicted values. Predicted values never use lags of other predicted values, as would be desirable for dynamic simulation. For these applications, PROC SIMLIN is available to predict or simulate values from the estimated equations.

OUTEST= Data Set

The OUTEST= option produces a TYPE=EST output SAS data set containing estimates from the regressions. The variables in the OUTEST= data set are as follows:

BY variables	the BY statement variables are included in the OUTEST= data set
TYPE	identifies the estimation type for the observations. The _TYPE_ value INST indicates first-stage regression estimates. Other values indicate the estimation method used: 2SLS indicates two-stage

least squares results, 3SLS indicates three-stage least squares results, LIML indicates limited information maximum likelihood results, and so forth. Observations added by IDENTITY statements have the _TYPE_ value IDENTITY.

MODEL the model label. The model label is the label specified on the MODEL statement or the dependent variable name if no label is specified. For first-stage regression estimates, _MODEL_ has the value FIRST.

DEPVAR the name of the dependent variable for the model

NAME the names of the regressors for the rows of the covariance matrix, if the COVOUT option is specified. _NAME_ has a blank value for the parameter estimates observations. The _NAME_ variable is not included in the OUTEST= data set unless the COVOUT option is used to output the covariance of parameter estimates matrix.

SIGMA contains the root mean square error for the model, which is an estimate of the standard deviation of the error term. The _SIGMA_ variable contains the same values reported as Root MSE in the printed output.

INTERCEPT the intercept parameter estimates

regressors the regressor variables from all the MODEL statements are included in the OUTEST= data set. Variables used in IDENTIFY statements are also included in the OUTEST= data set.

The parameter estimates are stored under the names of the regressor variables. The intercept parameters are stored in the variable INTERCEP. The dependent variable of the model is given a coefficient of -1. Variables not in a model have missing values for the OUTEST= observations for that model.

Some estimation methods require computation of preliminary estimates. All estimates computed are output to the OUTEST= data set. For each BY group and each estimation, the OUTEST= data set contains one observation for each MODEL or IDENTITY statement. Results for different estimations are identified by the _TYPE_ variable.

For example, consider the following statements:

```
proc syslin data=a outest=est 3sls;
   by b;
   endogenous y1 y2;
   instruments x1-x4;
   model y1 = y2 x1 x2;
   model y2 = y1 x3 x4;
   identity x1 = x3 + x4;
run;
```

The 3SLS method requires both a preliminary 2SLS stage and preliminary first stage regressions for the endogenous variable. The OUTEST= data set thus contains 3 different kinds of estimates. The observations for the first-stage regression estimates have the _TYPE_ value INST. The observations for the 2SLS estimates have the _TYPE_ value 2SLS. The observations for the final 3SLS estimates have the _TYPE_ value 3SLS.

Since there are 2 endogenous variables in this example, there are 2 first-stage regressions and 2 _TYPE_=INST observations in the OUTEST= data set. Since there are 2 model statements, there are 2 OUTEST= observations with _TYPE_=2SLS and 2 observations with _TYPE_=3SLS. In addition, the OUTEST= data set contains an observation with the _TYPE_ value IDENTITY containing the coefficients specified by the IDENTITY statement. All these observations are repeated for each BY-group in the input data set defined by the values of the BY variable B.

When the COVOUT option is specified, the estimated covariance matrix for the parameter estimates is included in the OUTEST= data set. Each observation for parameter estimates is followed by observations containing the rows of the parameter covariance matrix for that model. The row of the covariance matrix is identified by the variable _NAME_. For observations that contain parameter estimates, _NAME_ is blank. For covariance observations, _NAME_ contains the regressor name for the row of the covariance matrix, and the regressor variables contain the covariances.

See Example 19.1 for an example of the OUTEST= data set.

OUTSSCP= Data Set

The OUTSSCP= option produces a TYPE=SSCP output SAS data set containing sums of squares and cross products. The data set contains all variables used in the MODEL, IDENTITY, and VAR statements. Observations are identified by the variable _NAME_.

The OUTSSCP= data set can be useful when a large number of observations are to be explored in many different SYSLIN runs. The sum-of-squares-and-crossproducts matrix can be saved with the OUTSSCP= option and used as the DATA= data set on subsequent SYSLIN runs. This is much less expensive computationally because PROC SYSLIN never reads the original data again. In the step that creates the OUTSSCP= data set, include in the VAR statement all the variables you expect to use.

Printed Output

The printed output produced by the SYSLIN procedure is as follows:

1. If the SIMPLE option is used, a table of descriptive statistics is printed showing the sum, mean, sum of squares, variance, and standard deviation for all the variables used in the models.
2. First-stage regression results are printed if the FIRST option is specified and an instrumental variables method is used. This shows the regression of each endogenous variable on the variables in the INSTRUMENTS list.
3. The results of the second-stage regression are printed for each model. (See "Printed Output for Each Model," which follows.)

4. If a systems method like 3SLS, SUR, or FIML is used, the cross-equation error covariance matrix is printed. This matrix is shown four ways: the covariance matrix itself, the correlation matrix form, the inverse of the correlation matrix, and the inverse of the covariance matrix.

5. If a systems method like 3SLS, SUR, or FIML is used, the system weighted mean square error and system weighted R^2 statistics are printed. The system weighted MSE and R^2 measure the fit of the joint model obtained by stacking all the models together and performing a single regression with the stacked observations weighted by the inverse of the model error variances.

6. If a systems method like 3SLS, SUR, or FIML is used, the final results are printed for each model.

7. If the REDUCED option is used, the reduced form coefficients are printed. This consists of the structural coefficient matrix for the endogenous variables, the structural coefficient matrix for the exogenous variables, the inverse of the endogenous coefficient matrix, and the reduced form coefficient matrix. The reduced form coefficient matrix is the product of the inverse of the endogenous coefficient matrix and the exogenous structural coefficient matrix.

Printed Output for Each Model

The results printed for each model include the "Analysis of Variance" table, the "Parameter Estimates" table, and optional items requested by TEST statements or by options on the MODEL statement.

The printed output produced for each model is described in the following.

The Analysis of Variance table includes the following:

- the model degrees of freedom, sum of squares, and mean square
- the error degrees of freedom, sum of squares, and mean square. The error mean square is computed by dividing the error sum of squares by the error degrees of freedom and is not effected by the VARDEF= option.
- the corrected total degrees of freedom and total sum of squares. Note that for instrumental variables methods the model and error sums of squares do not add to the total sum of squares.
- the F-ratio, labeled "F Value," and its significance, labeled "PROB>F," for the test of the hypothesis that all the nonintercept parameters are 0
- the root mean square error. This is the square root of the error mean square.
- the dependent variable mean
- the coefficient of variation (C.V.) of the dependent variable
- the R^2 statistic. This R^2 is computed consistently with the calculation of the F statistic. It is valid for hypothesis tests but may not be a good measure of fit for models estimated by instrumental variables methods.
- the R^2 statistic adjusted for model degrees of freedom, labeled "Adj R-SQ"

The Parameter Estimates table includes the following.

- estimates of parameters for regressors in the model and the Lagrangian parameter for each restriction specified
- a degrees of freedom column labeled DF. Estimated model parameters have 1 degree of freedom. Restrictions have a DF of -1. Regressors or restrictions dropped from the model due to collinearity have a DF of 0.
- the standard errors of the parameter estimates
- the t statistics, which are the parameter estimates divided by the standard errors
- the significance of the *t*-tests for the hypothesis that the true parameter is 0, labeled "Pr > |t|." As previously noted, the significance tests are strictly valid in finite samples only for OLS estimates but are asymptotically valid for the other methods.
- the standardized regression coefficients, if the STB option is specified. This is the parameter estimate multiplied by the ratio of the standard deviation of the regressor to the standard deviation of the dependent variable.
- the labels of the regressor variables or restriction labels

In addition to the Analysis of Variance table and the Parameter Estimates table, the results printed for each model may include the following:

1. If TEST statements are specified, the test results are printed.
2. If the DW option is specified, the Durbin-Watson statistic and first-order autocorrelation coefficient are printed.
3. If the OVERID option is specified, the results of Basmann's test for overidentifying restrictions are printed.
4. If the PLOT option is used, plots of residual against each regressor are printed.
5. If the COVB or CORRB options are specified, the results for each model also include the covariance or correlation matrix of the parameter estimates. For systems methods like 3SLS and FIML, the COVB and CORB output is printed for the whole system after the output for the last model, instead of separately for each model.

The third stage output for 3SLS, SUR, IT3SLS, ITSUR, and FIML does not include the Analysis of Variance table. When a systems method is used, the second stage output does not include the optional output, except for the COVB and CORB matrices.

ODS Table Names

PROC SYSLIN assigns a name to each table it creates. You can use these names to reference the table when using the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in the following table. For more information on ODS, see Chapter 6, "Using the Output Delivery System."

Table 19.1. ODS Tables Produced in PROC SYSLIN

ODS Table Name	Description	Option
ANOVA	Summary of the SSE, MSE for the equations	default
AugXPXMat	Model Crossproducts	XPX
AutoCorrStat	Autocorrelation Statistics	default
ConvCrit	Convergence criteria for estimation	default
ConvergenceStatus	Convergence status	default
CorrB	Correlations of parameters	CORRB
CorrResiduals	Correlations of residuals	CORRS
CovB	Covariance of parameters	COVB
CovResiduals	Covariance of residuals	COVS
EndoMat	Endogenous Variables	
Equations	Listing of equations to estimate	default
ExogMat	Exogenous Variables	
FitStatistics	Statistics of Fit	default
InvCorrResiduals	Inverse Correlations of residuals	CORRS
InvCovResiduals	InvCovariance of residuals	COVS
InvEndoMat	Inverse Endogenous Variables	
InvXPX	$X'X$ inverse for System	I
IterHistory	Iteration printing	ITALL/ITPRINT
MissingValues	Missing values generated by the program	default
ModelVars	Name and label for the Model	default
ParameterEstimates	Parameter Estimates	default
RedMat	Reduced Form	REDUCED
SimpleStatistics	Descriptive statistics	SIMPLE
SSCP	Model Crossproducts	
TestResults	Test for Overidentifying Restrictions	
Weight	Weighted Model Statistics	
YPY	Y'Y matrices	USSCP2

Examples

Example 19.1. Klein’s Model I Estimated with LIML and 3SLS

This example uses PROC SYSLIN to estimate the classic Klein Model I. For a discussion of this model, see Theil (1971). The following statements read the data.

```
*-------------------Klein's Model I-------------------------*
| By L.R. Klein, Economic Fluctuations in the United States,  |
| 1921-1941 (1950), NY: John Wiley.   A macro-economic model  |
| of the U.S. with three behavioral equations, and several    |
| identities. See Theil, p.456.                               |
*-------------------------------------------------------------*;
data klein;
   input year c p w i x wp g t k wsum;
   date=mdy(1,1,year);
   format date monyy.;
   y   =c+i+g-t;
   yr  =year-1931;
   klag=lag(k);
   plag=lag(p);
   xlag=lag(x);
   label year='Year'
         date='Date'
         c   ='Consumption'
         p   ='Profits'
         w   ='Private Wage Bill'
         i   ='Investment'
         k   ='Capital Stock'
         y   ='National Income'
         x   ='Private Production'
         wsum='Total Wage Bill'
         wp  ='Govt Wage Bill'
         g   ='Govt Demand'
         i   ='Taxes'
         klag='Capital Stock Lagged'
         plag='Profits Lagged'
         xlag='Private Product Lagged'
         yr  ='YEAR-1931';
   datalines;
1920     .  12.7     .    .  44.9    .     .     .  182.8     .
1921  41.9  12.4  25.5 -0.2  45.6  2.7   3.9   7.7  182.6  28.2
1922  45.0  16.9  29.3  1.9  50.1  2.9   3.2   3.9  184.5  32.2
1923  49.2  18.4  34.1  5.2  57.2  2.9   2.8   4.7  189.7  37.0
1924  50.6  19.4  33.9  3.0  57.1  3.1   3.5   3.8  192.7  37.0
1925  52.6  20.1  35.4  5.1  61.0  3.2   3.3   5.5  197.8  38.6
1926  55.1  19.6  37.4  5.6  64.0  3.3   3.3   7.0  203.4  40.7
1927  56.2  19.8  37.9  4.2  64.4  3.6   4.0   6.7  207.6  41.5
1928  57.3  21.1  39.2  3.0  64.5  3.7   4.2   4.2  210.6  42.9
1929  57.8  21.7  41.3  5.1  67.0  4.0   4.1   4.0  215.7  45.3
1930  55.0  15.6  37.9  1.0  61.2  4.2   5.2   7.7  216.7  42.1
1931  50.9  11.4  34.5 -3.4  53.4  4.8   5.9   7.5  213.3  39.3
1932  45.6   7.0  29.0 -6.2  44.3  5.3   4.9   8.3  207.1  34.3
1933  46.5  11.2  28.5 -5.1  45.1  5.6   3.7   5.4  202.0  34.1
1934  48.7  12.3  30.6 -3.0  49.7  6.0   4.0   6.8  199.0  36.6
1935  51.3  14.0  33.2 -1.3  54.4  6.1   4.4   7.2  197.7  39.3
1936  57.7  17.6  36.8  2.1  62.7  7.4   2.9   8.3  199.8  44.2
```

```
1937  58.7  17.3  41.0  2.0  65.0  6.7   4.3   6.7  201.8  47.7
1938  57.5  15.3  38.2 -1.9  60.9  7.7   5.3   7.4  199.9  45.9
1939  61.6  19.0  41.6  1.3  69.5  7.8   6.6   8.9  201.2  49.4
1940  65.0  21.1  45.0  3.3  75.7  8.0   7.4   9.6  204.5  53.0
1941  69.7  23.5  53.3  4.9  88.4  8.5  13.8  11.6  209.4  61.8
;
```

The following statements estimate the Klein model using the limited information maximum likelihood method. In addition, the parameter estimates are written to a SAS data set with the OUTEST= option.

```
proc syslin data=klein outest=b liml;
   endogenous c p w i x wsum k y;
   instruments klag plag xlag wp g t yr;
   consume: model c = p plag  wsum;
   invest:  model i = p plag  klag;
   labor:   model w = x xlag  yr;
run;

proc print data=b; run;
```

The PROC SYSLIN estimates are shown in Output 19.1.1.

Output 19.1.1. LIML Estimates

```
                         The SYSLIN Procedure
            Limited-Information Maximum Likelihood Estimation

                    Model                     CONSUME
                    Dependent Variable              c
                    Label                 Consumption


                         Analysis of Variance

                                  Sum of        Mean
     Source               DF     Squares      Square    F Value    Pr > F

     Model                 3    854.3541    284.7847     118.42    <.0001
     Error                17    40.88419    2.404952
     Corrected Total      20    941.4295


             Root MSE            1.55079    R-Square       0.95433
             Dependent Mean     53.99524    Adj R-Sq       0.94627
             Coeff Var           2.87209


                         Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  17.14765 2.045374    8.38   <.0001 Intercept
p              1  -0.22251 0.224230   -0.99   0.3349 Profits
plag           1  0.396027 0.192943    2.05   0.0558 Profits Lagged
wsum           1  0.822559 0.061549   13.36   <.0001 Total Wage Bill
```

```
                          The SYSLIN Procedure
               Limited-Information Maximum Likelihood Estimation

                         Model                   INVEST
                         Dependent Variable           i
                         Label                    Taxes


                          Analysis of Variance

                                      Sum of        Mean
     Source                 DF       Squares      Square    F Value    Pr > F

     Model                   3      210.3790    70.12634      34.06    <.0001
     Error                  17      34.99649    2.058617
     Corrected Total        20      252.3267


             Root MSE             1.43479    R-Square       0.85738
             Dependent Mean       1.26667    Adj R-Sq       0.83221
             Coeff Var          113.27274


                          Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  22.59083 9.498146    2.38   0.0294 Intercept
p              1  0.075185 0.224712    0.33   0.7420 Profits
plag           1  0.680386 0.209145    3.25   0.0047 Profits Lagged
klag           1  -0.16826 0.045345   -3.71   0.0017 Capital Stock Lagged
```

```
                              The SYSLIN Procedure
               Limited-Information Maximum Likelihood Estimation

                    Model                               LABOR
                    Dependent Variable                      w
                    Label                  Private Wage Bill


                              Analysis of Variance

                                    Sum of        Mean
     Source                 DF     Squares      Square    F Value    Pr > F

     Model                   3    696.1485    232.0495     393.62    <.0001
     Error                  17    10.02192    0.589525
     Corrected Total        20    794.9095


             Root MSE             0.76781    R-Square       0.98581
             Dependent Mean      36.36190    Adj R-Sq       0.98330
             Coeff Var            2.11156


                              Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  1.526187 1.320838    1.16   0.2639 Intercept
x              1  0.433941 0.075507    5.75   <.0001 Private Production
xlag           1  0.151321 0.074527    2.03   0.0583 Private Product Lagged
yr             1  0.131593 0.035995    3.66   0.0020 YEAR-1931
```

The OUTEST= data set is shown in part in Output 19.1.2. Note that the data set contains the parameter estimates and root mean square errors, _SIGMA_, for the first stage instrumental regressions as well as the parameter estimates and σ for the LIML estimates for the three structural equations.

Output 19.1.2. The OUTEST= Data Set

```
 Obs _TYPE_  _STATUS_   _MODEL_ _DEPVAR_ _SIGMA_ Intercept   klag     plag

  1   LIML  0 Converged CONSUME    c     1.55079  17.1477    .      0.39603
  2   LIML  0 Converged INVEST     i     1.43479  22.5908  -0.16826 0.68039
  3   LIML  0 Converged LABOR      w     0.76781   1.5262    .        .

 Obs   xlag   wp  g  t     yr     c      p      w   i     x       wsum   k  y

  1   .        .  .  .   .       -1  -0.22251   .   .   .       0.82256  .  .
  2   .        .  .  .   .        .   0.07518   .  -1   .        .       .  .
  3  0.15132   .  .  .  0.13159   .    .       -1   .  0.43394   .       .  .
```

The following statements estimate the model using the 3SLS method. The reduced form estimates are produced by the REDUCED option; IDENTITY statements are used to make the model complete.

```
proc syslin data=klein 3sls reduced;
   endogenous c p w i x wsum k y;
   instruments klag plag xlag wp g t yr;
   consume: model    c = p plag wsum;
```

```
      invest:  model    i = p plag klag;
      labor:   model    w = x xlag yr;
      product: identity x = c + i + g;
      income:  identity y = c + i + g - t;
      profit:  identity p = y - w;
      stock:   identity k = klag + i;
      wage:    identity wsum = w + wp;
   run;
```

The preliminary 2SLS results and estimated cross-model covariance matrix are not shown. The 3SLS estimates are shown in Output 19.1.3. The reduced form estimates are shown in Output 19.1.4.

Output 19.1.3. 3SLS Estimates

```
                              The SYSLIN Procedure
                      Three-Stage Least Squares Estimation

                     System Weighted MSE            5.9342
                     Degrees of freedom                 51
                     System Weighted R-Square       0.9550


                       Model                     CONSUME
                       Dependent Variable              c
                       Label                 Consumption


                              Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  16.44079 1.449925   11.34   <.0001 Intercept
p              1  0.124890 0.120179    1.04   0.3133 Profits
plag           1  0.163144 0.111631    1.46   0.1621 Profits Lagged
wsum           1  0.790081 0.042166   18.74   <.0001 Total Wage Bill
```

```
                              The SYSLIN Procedure
                      Three-Stage Least Squares Estimation

                         Model                   INVEST
                         Dependent Variable           i
                         Label                    Taxes


                              Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  28.17785 7.550853    3.73   0.0017 Intercept
p              1  -0.01308 0.179938   -0.07   0.9429 Profits
plag           1  0.755724 0.169976    4.45   0.0004 Profits Lagged
klag           1  -0.19485 0.036156   -5.39   <.0001 Capital Stock Lagged
```

```
                      The SYSLIN Procedure
               Three-Stage Least Squares Estimation

             Model                              LABOR
             Dependent Variable                     w
             Label                  Private Wage Bill


                      Parameter Estimates

              Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  1.797218 1.240203    1.45   0.1655 Intercept
x              1  0.400492 0.035359   11.33   <.0001 Private Production
xlag           1  0.181291 0.037965    4.78   0.0002 Private Product Lagged
yr             1  0.149674 0.031048    4.82   0.0002 YEAR-1931
```

Output 19.1.4. Reduced Form Estimates

```
                      The SYSLIN Procedure
               Three-Stage Least Squares Estimation

                      Endogenous Variables

                     c            p            w            i

        CONSUME      1     -0.12489            0            0
        INVEST       0     0.013079            0            1
        LABOR        0            0            1            0
        PRODUCT     -1            0            0           -1
        INCOME      -1            0            0           -1
        PROFIT       0            1            1            0
        STOCK        0            0            0           -1
        WAGE         0            0           -1            0

                      Endogenous Variables

                     x         wsum            k            y

        CONSUME      0     -0.79008            0            0
        INVEST       0            0            0            0
        LABOR -0.40049            0            0            0
        PRODUCT      1            0            0            0
        INCOME       0            0            0            1
        PROFIT       0            0            0           -1
        STOCK        0            0            1            0
        WAGE         0            1            0            0
```

The SYSLIN Procedure
Three-Stage Least Squares Estimation

Exogenous Variables

	Intercept	plag	klag	xlag
CONSUME	16.44079	0.163144	0	0
INVEST	28.17785	0.755724	-0.19485	0
LABOR	1.797218	0	0	0.181291
PRODUCT	0	0	0	0
INCOME	0	0	0	0
PROFIT	0	0	0	0
STOCK	0	0	1	0
WAGE	0	0	0	0

Exogenous Variables

	yr	g	t	wp
CONSUME	0	0	0	0
INVEST	0	0	0	0
LABOR	0.149674	0	0	0
PRODUCT	0	1	0	0
INCOME	0	1	-1	0
PROFIT	0	0	0	0
STOCK	0	0	0	0
WAGE	0	0	0	1

The SYSLIN Procedure
Three-Stage Least Squares Estimation

Inverse Endogenous Variables

	CONSUME	INVEST	LABOR	PRODUCT
c	1.634654	0.634654	1.095657	0.438802
p	0.972364	0.972364	-0.34048	-0.13636
w	0.649572	0.649572	1.440585	0.576943
i	-0.01272	0.987282	0.004453	0.001783
x	1.621936	1.621936	1.10011	1.440585
wsum	0.649572	0.649572	1.440585	0.576943
k	-0.01272	0.987282	0.004453	0.001783
y	1.621936	1.621936	1.10011	0.440585

Inverse Endogenous Variables

	INCOME	PROFIT	STOCK	WAGE
c	0.195852	0.195852	0	1.291509
p	1.108721	1.108721	0	0.768246
w	0.072629	0.072629	0	0.513215
i	-0.0145	-0.0145	0	-0.01005
x	0.181351	0.181351	0	1.281461
wsum	0.072629	0.072629	0	1.513215
k	-0.0145	-0.0145	1	-0.01005
y	1.181351	0.181351	0	1.281461

```
                    The SYSLIN Procedure
             Three-Stage Least Squares Estimation

                        Reduced Form

          Intercept          plag          klag          xlag

c           46.7273      0.746307      -0.12366      0.198633
p          42.77363      0.893474      -0.18946      -0.06173
w          31.57207      0.596871      -0.12657      0.261165
i           27.6184      0.744038      -0.19237      0.000807
x           74.3457      1.490345      -0.31603       0.19944
wsum       31.57207      0.596871      -0.12657      0.261165
k           27.6184      0.744038       0.80763      0.000807
y           74.3457      1.490345      -0.31603       0.19944

                        Reduced Form

                 yr             g             t            wp

 c         0.163991      0.634654      -0.19585      1.291509
 p         -0.05096      0.972364      -1.10872      0.768246
 w         0.215618      0.649572      -0.07263      0.513215
 i         0.000667      -0.01272      0.014501      -0.01005
 x         0.164658      1.621936      -0.18135      1.281461
 wsum      0.215618      0.649572      -0.07263      1.513215
 k         0.000667      -0.01272      0.014501      -0.01005
 y         0.164658      1.621936      -1.18135      1.281461
```

Example 19.2. Grunfeld’s Model Estimated with SUR

The following example was used by Zellner in his classic 1962 paper on seemingly unrelated regressions. Different stock prices often move in the same direction at a given point in time. The SUR technique may provide more efficient estimates than OLS in this situation.

The following statements read the data. (The prefix GE stands for General Electric and WH stands for Westinghouse.)

```
*---------Zellner’s Seemingly Unrelated Technique------------*
| A. Zellner, "An Efficient Method of Estimating Seemingly   |
| Unrelated Regressions and Tests for Aggregation Bias,"     |
| JASA 57(1962) pp.348-364                                   |
|                                                            |
| J.C.G. Boot, "Investment Demand: an Empirical Contribution |
| to the Aggregation Problem," IER 1(1960) pp.3-30.          |
|                                                            |
| Y. Grunfeld, "The Determinants of Corporate Investment,"   |
| Unpublished thesis, Chicago, 1958                          |
*------------------------------------------------------------*;

data grunfeld;
   input year ge_i ge_f ge_c wh_i wh_f wh_c;
   label ge_i = ’Gross Investment, GE’
         ge_c = ’Capital Stock Lagged, GE’
         ge_f = ’Value of Outstanding Shares Lagged, GE’
         wh_i = ’Gross Investment, WH’
         wh_c = ’Capital Stock Lagged, WH’
         wh_f = ’Value of Outstanding Shares Lagged, WH’;
```

```
   datalines;
1935     33.1      1170.6    97.8      12.93     191.5     1.8
1936     45.0      2015.8    104.4     25.90     516.0     .8
1937     77.2      2803.3    118.0     35.05     729.0     7.4
1938     44.6      2039.7    156.2     22.89     560.4     18.1
1939     48.1      2256.2    172.6     18.84     519.9     23.5
1940     74.4      2132.2    186.6     28.57     628.5     26.5
1941     113.0     1834.1    220.9     48.51     537.1     36.2
1942     91.9      1588.0    287.8     43.34     561.2     60.8
1943     61.3      1749.4    319.9     37.02     617.2     84.4
1944     56.8      1687.2    321.3     37.81     626.7     91.2
1945     93.6      2007.7    319.6     39.27     737.2     92.4
1946     159.9     2208.3    346.0     53.46     760.5     86.0
1947     147.2     1656.7    456.4     55.56     581.4     111.1
1948     146.3     1604.4    543.4     49.56     662.3     130.6
1949     98.3      1431.8    618.3     32.04     583.8     141.8
1950     93.5      1610.5    647.4     32.24     635.2     136.7
1951     135.2     1819.4    671.3     54.38     723.8     129.7
1952     157.3     2079.7    726.1     71.78     864.1     145.5
1953     179.5     2371.6    800.3     90.08     1193.5    174.8
1954     189.6     2759.9    888.9     68.60     1188.9    213.5
;
```

The following statements compute the SUR estimates for the Grunfeld model.

```
proc syslin data=grunfeld sur;
   ge:      model ge_i = ge_f ge_c;
   westing: model wh_i = wh_f wh_c;
run;
```

The PROC SYSLIN output is shown in Output 19.2.1.

Output 19.2.1. PROC SYSLIN Output for SUR

```
                          The SYSLIN Procedure
                    Ordinary Least Squares Estimation

                Model                                  GE
                Dependent Variable                   ge_i
                Label                 Gross Investment, GE


                          Analysis of Variance

                                     Sum of        Mean
     Source                 DF      Squares      Square    F Value    Pr > F

     Model                   2     31632.03    15816.02      20.34    <.0001
     Error                  17     13216.59    777.4463
     Corrected Total        19     44848.62


             Root MSE            27.88272    R-Square       0.70531
             Dependent Mean     102.29000    Adj R-Sq       0.67064
             Coeff Var           27.25850


                          Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  -9.95631 31.37425   -0.32   0.7548 Intercept
ge_f           1  0.026551 0.015566    1.71   0.1063 Value of Outstanding Shares
                                                     Lagged, GE
ge_c           1  0.151694 0.025704    5.90   <.0001 Capital Stock Lagged, GE
```

```
                          The SYSLIN Procedure
                   Ordinary Least Squares Estimation

               Model                              WESTING
               Dependent Variable                    wh_i
               Label                 Gross Investment, WH


                          Analysis of Variance

                                   Sum of        Mean
     Source                 DF    Squares      Square    F Value    Pr > F

     Model                   2   5165.553    2582.776      24.76    <.0001
     Error                  17   1773.234    104.3079
     Corrected Total        19   6938.787


             Root MSE           10.21312    R-Square       0.74445
             Dependent Mean     42.89150    Adj R-Sq       0.71438
             Coeff Var          23.81153


                          Parameter Estimates

                 Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  -0.50939 8.015289   -0.06   0.9501 Intercept
wh_f           1  0.052894 0.015707    3.37   0.0037 Value of Outstanding Shares
                                                     Lagged, WH
wh_c           1  0.092406 0.056099    1.65   0.1179 Capital Stock Lagged, WH
```

```
                         The SYSLIN Procedure
                Seemingly Unrelated Regression Estimation

                        Cross Model Covariance

                               GE            WESTING

                 GE             777.446       207.587
                 WESTING        207.587       104.308


                        Cross Model Correlation

                               GE            WESTING

                 GE             1.00000       0.72896
                 WESTING        0.72896       1.00000


                    Cross Model Inverse Correlation

                               GE            WESTING

                 GE             2.13397      -1.55559
                 WESTING       -1.55559       2.13397


                     Cross Model Inverse Covariance

                               GE            WESTING

                 GE            0.002745      -.005463
                 WESTING       -.005463      0.020458
```

```
                         The SYSLIN Procedure
                Seemingly Unrelated Regression Estimation

                 System Weighted MSE            0.9719
                 Degrees of freedom                 34
                 System Weighted R-Square       0.6284


               Model                                    GE
               Dependent Variable                     ge_i
               Label                   Gross Investment, GE


                          Parameter Estimates

             Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  -27.7193 29.32122   -0.95   0.3577 Intercept
ge_f           1  0.038310 0.014415    2.66   0.0166 Value of Outstanding Shares
                                                     Lagged, GE
ge_c           1  0.139036 0.024986    5.56   <.0001 Capital Stock Lagged, GE
```

```
                          The SYSLIN Procedure
               Seemingly Unrelated Regression Estimation

               Model                              WESTING
               Dependent Variable                    wh_i
               Label                 Gross Investment, WH


                          Parameter Estimates

             Parameter Standard                  Variable
Variable      DF  Estimate    Error t Value Pr > |t| Label

Intercept      1  -1.25199 7.545217   -0.17   0.8702 Intercept
wh_f           1  0.057630 0.014546    3.96   0.0010 Value of Outstanding Shares
                                                     Lagged, WH
wh_c           1  0.063978 0.053041    1.21   0.2443 Capital Stock Lagged, WH
```

References

Basmann, R.L. (1960), "On Finite Sample Distributions of Generalized Classical Linear Identifiability Test Statistics," *Journal of the American Statistical Association*, 55, 650-659.

Fuller, W.A. (1977), "Some Properties of a Modification of the Limited Information Estimator," *Econometrica*, 45, 939-952.

Hausman, J.A. (1975), "An Instrumental Variable Approach to Full Information Estimators for Linear and Certain Nonlinear Econometric Models," *Econometrica*, 43, 727-738.

Johnston, J. (1984), *Econometric Methods*, Third Edition, New York: McGraw-Hill Book Company.

Judge, George G., W. E. Griffiths, R. Carter Hill, Helmut Lutkepohl, and Tsoung-Chao Lee (1985), *The Theory and Practice of Econometrics*, Second Edition, New York: John Wiley & Sons, Inc.

Maddala, G.S. (1977), *Econometrics*, New York: McGraw-Hill Book Company.

Park, S.B. (1982), "Some Sampling Properties of Minimum Expected Loss (MELO) Estimators of Structural Coefficients," *Journal of the Econometrics*, 18, 295-311.

Pindyck, R.S. and Rubinfeld, D.L. (1981), *Econometric Models and Economic Forecasts*, Second Edition, New York: McGraw-Hill Book Company.

Pringle, R.M. and Raynor, A.A. (1971), *Generalized Inverse Matrices with Applications to Statistics*, New York: Hafner Publishing Company.

Rao, P. (1974), "Specification Bias in Seemingly Unrelated Regressions," in *Essays in Honor of Tinbergen*, Volume 2, New York: International Arts and Sciences Press.

Savin, N.E. and White, K.J. (1978), "Testing for Autocorrelation with Missing Observations," *Econometrics*, 46, 59-66.

Theil, H. (1971), *Principles of Econometrics*, New York: John Wiley & Sons, Inc.

Zellner, A. (1962), "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association*, 57, 348-368.

Zellner, A. (1978), "Estimation of Functions of Population Means and Regression Coefficients: A Minimum Expected Loss (MELO) Approach," *Journal of the Econometrics*, 8, 127-158.

Zellner, A. and Park, S. (1979), "Minimum Expected Loss (MELO) Estimators for Functions of Parameters and Structural Coefficients of Econometric Models," *Journal of the American Statistical Association*, 74, 185-193.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *SAS/ETS User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999. 1546 pp.

SAS/ETS User's Guide, Version 8

ISBN 1–58025–489–6

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, October 1999