Chapter 20
The TSCSREG Procedure

Chapter Table of Contents

Chapter 20
The TSCSREG Procedure

Overview

The TSCSREG (**T**ime **S**eries **C**ross **S**ection **Reg**ression) procedure analyzes a class of linear econometric models that commonly arise when time series and cross-sectional data are combined. The TSCSREG procedure deals with panel data sets that consist of time series observations on each of several cross-sectional units.

Such models can be viewed as two-way designs with covariates

$$y_{it} = \sum_{k=1}^{K} X_{itk}\beta_k + u_{it} \quad i = 1, \ldots, N; \quad t = 1, \ldots, T$$

where *N* is the number of cross sections, *T* is the length of the time series for each cross section, and *K* is the number of exogenous or independent variables.

The performance of any estimation procedure for the model regression parameters depends on the statistical characteristics of the error components in the model. The TSCSREG procedure estimates the regression parameters in the preceding model under several common error structures. The error structures and the corresponding methods the TSCSREG procedure uses to analyze them are as follows:

- one and two-way fixed and random effects models. If the specification is dependent only on the cross section to which the observation belongs, such a model is referred to as a model with one-way effects. A specification that depends on both the cross section and the time series to which the observation belongs is called a model with two-way effects.
- Therefore, the specifications for the one-way model are

  $$u_{it} = \nu_i + \epsilon_{it}$$

  and the specifications for the two-way model are

  $$u_{it} = \nu_i + e_t + \epsilon_{it}$$

  where ϵ_{it} is a classical error term with zero mean and a homoscedastic covariance matrix.
- Apart from the possible one-way or two-way nature of the effect, the other dimension of difference between the possible specifications is that of the nature of the cross-sectional or time-series effect. The models are referred to as fixed effects models if the effects are nonrandom and as random effects models otherwise.

- first-order autoregressive model with contemporaneous correlation

$$u_{it} = \rho_i u_{i,t-1} + \epsilon_{it}$$

- The Parks method is used to estimate this model. This model assumes a first-order autoregressive error structure with contemporaneous correlation between cross sections. The covariance matrix is estimated by a two-stage procedure leading to the estimation of model regression parameters by GLS.
- mixed variance-component moving average error process

$$u_{it} = a_i + b_t + e_{it}$$

$$e_{it} = \alpha_0 \epsilon_t + \alpha_1 \epsilon_{t-1} + \ldots + \alpha_m \epsilon_{t-m}$$

- The Da Silva method is used to estimate this model. The Da Silva method estimates the regression parameters using a two-step GLS-type estimator.

The TSCSREG procedure analyzes panel data sets that consist of multiple time series observations on each of several individuals or cross-sectional units. The input data set must be in time series cross-sectional form. See Chapter 2, “Working with Time Series Data,” for a discussion of how time series related by a cross-sectional dimension are stored in SAS data sets. The TSCSREG procedure requires that the time series for each cross section have the same number of observations and cover the same time range.

Getting Started

Specifying the Input Data

The input data set used by the TSCSREG procedure must be sorted by cross section and by time within each cross section. Therefore, the first step in using PROC TSCSREG is to make sure that the input data set is sorted. Normally, the input data set contains a variable that identifies the cross section for each observation and a variable that identifies the time period for each observation.

To illustrate, suppose that you have a data set A containing data over time for each of several states. You want to regress the variable Y on regressors X1 and X2. Cross sections are identified by the variable STATE, and time periods are identified by the variable DATE. The following statements sort the data set A appropriately:

```
proc sort data=a;
   by state date;
run;
```

The next step is to invoke the TSCSREG procedure and specify the cross section and time series variables in an ID statement. List the variables in the ID statement exactly as they are listed in the BY statement.

```
proc tscsreg data=a;
   id state date;
```

Alternatively, you can omit the ID statement and use the CS= and TS= options on the PROC TSCSREG statement to specify the number of cross sections in the data set and the number of time series observations in each cross section.

Unbalanced Data

In the case of fixed effects and random effects models, the TSCSREG procedure is capable of processing data with different numbers of time series observations across different cross sections. You must specify the ID statement to estimate models using unbalanced data. The missing time series observations are recognized by the absence of time series id variable values in some of the cross sections in the input data set. Moreover, if an observation with a particular time series id value and cross-sectional id value is present in the input data set, but one or more of the model variables are missing, that time series point is treated as missing for that cross section.

Also, when PROC TSCSREG is processing balanced data, you now need to specify only the CS= parameter if you do not specify an ID statement. The TS= parameter is not required, since it can be inferred from the number of observations if the data is balanced.

Specifying the Regression Model

Next, specify the linear regression model with a MODEL statement. The MODEL statement in PROC TSCSREG is specified like the MODEL statement in other SAS regression procedures: the dependent variable is listed first, followed by an equal sign, followed by the list of regressor variables.

```
proc tscsreg data=a;
   id state date;
   model y = x1 x2;
run;
```

The reason for using PROC TSCSREG instead of other SAS regression procedures is that you can incorporate a model for the structure of the random errors. It is important to consider what kind of error structure model is appropriate for your data and to specify the corresponding option in the MODEL statement.

The error structure options supported by the TSCSREG procedure are FIXONE, FIXTWO, RANONE, RANTWO, FULLER, PARKS, and DASILVA. See the "Details" section later in this chapter for more information about these methods and the error structures they assume.

By default, the Fuller-Battese method is used. Thus, the preceding example is the same as specifying the FULLER option, as shown in the following statements:

```
proc tscsreg data=a;
   id state date;
   model y = x1 x2 / fuller;
run;
```

You can specify more than one error structure option in the MODEL statement; the analysis is repeated using each method specified. You can use any number of MODEL statements to estimate different regression models or estimate the same model using different options. See Example 20.1 in the section "Examples."

In order to aid in model specification within this class of models, the procedure provides two specification test statistics. The first is an F statistic that tests the null hypothesis that the fixed effects parameters are all zero. The second is a Hausman m-statistic that provides information about the appropriateness of the random effects specification. It is based on the idea that, under the null hypothesis of no correlation between the effects variables and the regressors, OLS and GLS are consistent, but OLS is inefficient. Hence, a test can be based on the result that the covariance of an efficient estimator with its difference from an inefficient estimator is zero. Rejection of the null hypothesis might suggest that the fixed effects model is more appropriate.

The procedure also provides the Buse R-squared measure, which is the most appropriate goodness-of-fit measure for models estimated using GLS. This number is interpreted as a measure of the proportion of the transformed sum of squares of the dependent variable that is attributable to the influence of the independent variables. In the case of OLS estimation, the Buse R-squared measure is equivalent to the usual R-squared measure.

Estimation Techniques

If the effects are fixed, the models are essentially regression models with dummy variables corresponding to the specified effects. For fixed effects models, ordinary least squares (OLS) estimation is best linear unbiased.

The other alternative is to assume that the effects are random. In the one-way case, $E(\nu_i) = 0$, $E(\nu_i^2) = \sigma_\nu^2$, and

$E(\nu_i \nu_j) = 0$ for $i \neq j$, and ν_i is uncorrelated with ϵ_{it} for all i and t. In the two-way case, in addition to all of the preceding, $E(e_t) = 0$, $E(e_t^2) = \sigma_e^2$, and

$E(e_t e_s) = 0$ for $t \neq s$, and the e_t are uncorrelated with the ν_i and the ϵ_{it} for all $i$and t. Thus, the model is a variance components model, with the variance components σ_ν^2 and σ_e^2, as well as σ_ϵ^2, to be estimated. A crucial implication of such a specification is that the effects are independent of the regressors. For random effects models, the estimation method is an estimated generalized least squares (EGLS) procedure that involves estimating the variance components in the first stage and using the estimated variance covariance matrix thus obtained to apply generalized least squares (GLS) to the data.

Introductory Example

The following example uses the cost function data from Greene (1990) to estimate the variance components model. The variable OUTPUT is the log of output in millions of kilowatt-hours, and COST is the log of cost in millions of dollars. Refer to Greene (1990) for details.

```
data greene;
   input firm year output cost @@;
cards;
     1 1955   5.36598   1.14867  1 1960   6.03787   1.45185
     1 1965   6.37673   1.52257  1 1970   6.93245   1.76627
     2 1955   6.54535   1.35041  2 1960   6.69827   1.71109
     2 1965   7.40245   2.09519  2 1970   7.82644   2.39480
     3 1955   8.07153   2.94628  3 1960   8.47679   3.25967
     3 1965   8.66923   3.47952  3 1970   9.13508   3.71795
     4 1955   8.64259   3.56187  4 1960   8.93748   3.93400
     4 1965   9.23073   4.11161  4 1970   9.52530   4.35523
     5 1955   8.69951   3.50116  5 1960   9.01457   3.68998
     5 1965   9.04594   3.76410  5 1970   9.21074   4.05573
     6 1955   9.37552   4.29114  6 1960   9.65188   4.59356
     6 1965  10.21163   4.93361  6 1970  10.34039   5.25520
;

proc sort data=greene;
   by firm year;
run;
```

Usually you cannot explicitly specify all the explanatory variables that affect the dependent variable. The omitted or unobservable variables are summarized in the error disturbances. The TSCSREG procedure used with the Fuller-Battese method adds the individual and time-specific random effects to the error disturbances, and the parameters are efficiently estimated using the GLS method. The variance components model used by the Fuller-Battese method is

$$y_{it} = \sum_{k=1}^{K} X_{itk}\beta_k + v_i + e_t + \epsilon_{it} \quad i = 1, \ldots, N; \quad t = 1, \ldots, T$$

The following statements fit this model. Since the Fuller-Battese is the default method, no options are required.

```
proc tscsreg data=greene;
   model cost = output;
   id firm year;
run;
```

The TSCSREG procedure output is shown in Figure 20.1. A model description is printed first, which reports the estimation method used and the number of cross sections and time periods. The variance components estimates are printed next. Finally, the table of regression parameter estimates shows the estimates, standard errors, and *t*-tests.

```
                          The TSCSREG Procedure


Dependent Variable: cost

                            Model Description

                    Estimation Method              RanTwo
                    Number of Cross Sections            6
                    Time Series Length                  4


                              Fit Statistics

             SSE              0.3481    DFE                  22
             MSE              0.0158    Root MSE         0.1258
             R-Square         0.8136


                      Variance Component Estimates

             Variance Component for Cross Sections    0.046907
             Variance Component for Time Series        0.00906
             Variance Component for Error             0.008749


                              Hausman Test for
                               Random Effects

                            DF    m Value    Pr > m

                             1      26.46    <.0001


                           Parameter Estimates

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

      Intercept        1     -2.99992         0.6478      -4.63      0.0001
      output           1     0.746596         0.0762       9.80      <.0001
```

Figure 20.1. The Variance Components Estimates

Syntax

The following statements are used with the TSCSREG procedure.

PROC TSCSREG *options***;**
BY *variables***;**
ID *cross-section-id-variable time-series-id-variable***;**
MODEL *dependent = regressor-variables / options***;**
label: **TEST** *equation [,equation...]***;**

Functional Summary

The statements and options used with the TSCSREG procedure are summarized in the following table.

Description	Statement	Option
Data Set Options		
specify the input data set	TSCSREG	DATA=
write parameter estimates to an output data set	TSCSREG	OUTEST=
include correlations in the OUTEST= data set	TSCSREG	CORROUT
include covariances in the OUTEST= data set	TSCSREG	COVOUT
specify number of time series observations	TSCSREG	TS=
specify number of cross sections	TSCSREG	CS=
Declaring the Role of Variables		
specify BY-group processing	BY	
specify the cross section and time ID variables	ID	
Printing Control Options		
print correlations of the estimates	MODEL	CORRB
print covariances of the estimates	MODEL	COVB
suppress printed output	MODEL	NOPRINT
perform tests of linear hypotheses	TEST	
Model Estimation Options		
specify the one-way fixed effects model	MODEL	FIXONE
specify the two-way fixed effects model	MODEL	FIXTWO
specify the one-way random effects model	MODEL	RANONE
specify the one-way random effects model	MODEL	RANTWO
specify Fuller-Battese method	MODEL	FULLER
specify PARKS	MODEL	PARKS
specify Da Silva method	MODEL	DASILVA
specify order of the moving average error process for Da Silva method	MODEL	M=

Description	Statement	Option
print Φ matrix for Parks method	MODEL	PHI
print autocorrelation coefficients for Parks method	MODEL	RHO
suppress the intercept term	MODEL	NOINT
control check for singularity	MODEL	SINGULAR=

PROC TSCSREG Statement

PROC TSCSREG *options;*

The following options can be specified on the PROC TSCSREG statement.

DATA= *SAS-data-set*
names the input data set. The input data set must be sorted by cross section and by time period within cross section. If you omit DATA=, the most recently created SAS data set is used.

TS= *number*
specifies the number of observations in the time series for each cross section. The TS= option value must be greater than 1. The TS= option is required unless an ID statement is used. Note that the number of observations for each time series must be the same for each cross section and must cover the same time period.

CS= *number*
specifies the number of cross sections. The CS= option value must be greater than 1. The CS= option is required unless an ID statement is used.

OUTEST= *SAS-data-set*
names an output data set to contain the parameter estimates. When the OUTEST= option is not specified, the OUTEST= data set is not created. See the section "OUTEST= Data Set" later in this chapter for details on the structure of the OUTEST= data set.

OUTCOV
COVOUT
writes the covariance matrix of the parameter estimates to the OUTEST= data set. See the section "OUTEST= Data Set" later in this chapter for details.

OUTCORR
CORROUT
writes the correlation matrix of the parameter estimates to the OUTEST= data set. See the section "OUTEST= Data Set" later in this chapter for details.

In addition, any of the following MODEL statement options can be specified in the PROC TSCSREG statement: CORRB, COVB, FIXONE, FIXTWO, RANONE, RANTWO, FULLER, PARKS, DASILVA, NOINT, NOPRINT, M=, PHI, RHO, and

SINGULAR=. When specified in the PROC TSCSREG statement, these options are equivalent to specifying the options for every MODEL statement. See the section "MODEL Statement" for a complete description of each of these options.

BY Statement

BY *variables ;*

A BY statement can be used with PROC TSCSREG to obtain separate analyses on observations in groups defined by the BY variables. When a BY statement appears, the input data set must be sorted by the BY variables as well as by cross section and time period within the BY groups.

When both an ID statement and a BY statement are specified, the input data set must be sorted first with respect to BY variables and then with respect to the cross section and time series ID variables. For example,

```
proc sort data=a;
   by byvar1 byvar2 csid tsid;
run;

proc tscsreg data=a;
   by byvar1 byvar2;
   id csid tsid;
   ...
run;
```

When both a BY statement and an ID statement are used, the data set may have a different number of cross sections or a different number of time periods in each BY group. If no ID statement is used, the CS=N and TS=T options must be specified and each BY group must contain $N \times T$ observations.

ID Statement

ID *cross-section-id-variable time-series-id-variable;*

The ID statement is used to specify variables in the input data set that identify the cross section and time period for each observation.

When an ID statement is used, the TSCSREG procedure verifies that the input data set is sorted by the cross section ID variable and by the time series ID variable within each cross section. The TSCSREG procedure also verifies that the time series ID values are the same for all cross sections.

To make sure the input data set is correctly sorted, use PROC SORT with a BY statement with the variables listed exactly as they are listed in the ID statement to sort the input data set.

```
proc sort data=a;
   by csid tsid;
run;

proc tscsreg data=a;
   id csid tsid;
   ... etc. ...
run;
```

If the ID statement is not used, the TS= and CS= options must be specified on the PROC TSCSREG statement. Note that the input data must be sorted by time within cross section, regardless of whether the cross section structure is given by an ID statement or by the options TS= and CS=.

If an ID statement is specified, the time series length T is set to the minimum number of observations for any cross section, and only the first T observations in each cross section are used. If both the ID statement and the TS= and CS= options are specified, the TS= and CS= options are ignored.

MODEL Statement

MODEL *response = regressors / options;*

The MODEL statement specifies the regression model and the error structure assumed for the regression residuals. The response variable on the left side of the equal sign is regressed on the independent variables listed after the equal sign. Any number of MODEL statements can be used. For each model statement only one response variable can be specified on the left side of the equal sign.

The error structure is specified by the FULLER, PARKS, and DASILVA options. More than one of these three options can be used, in which case the analysis is repeated for each error structure model specified.

Models can be given labels. Model labels are used in the printed output to identify the results for 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 specified on the MODEL statement after a slash (/).

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

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

FIXONE
specifies that a one-way fixed effects model be estimated.

FIXTWO
specifies that a two-way fixed effects model be estimated.

RANONE
specifies that a one-way random effects model be estimated.

RANTWO
specifies that a two-way random effects model be estimated.

FULLER
specifies that the model be estimated using the Fuller-Battese method, which assumes a variance components model for the error structure. See "Fuller-Battese Method" later in this chapter for details. FULLER is the default.

PARKS
specifies that the model be estimated using the Parks method, which assumes a first-order autoregressive model for the error structure. See "Parks Method" later in this chapter for details.

DASILVA
specifies that the model be estimated using the Da Silva method, which assumes a mixed variance-component moving average model for the error structure. See "Da Silva Method" later in this chapter for details.

M= *number*
specifies the order of the moving average process in the Da Silva method. The M= value must be less than $T - 1$. The default is M=1.

PHI
prints the Φ matrix of estimated covariances of the observations for the Parks method. The PHI option is relevant only when the PARKS option is used. See "Parks Method" later in this chapter for details.

RHO
prints the estimated autocorrelation coefficients for the Parks method.

NOINT
NOMEAN
suppresses the intercept parameter from the model.

NOPRINT
suppresses the normal printed output.

SINGULAR= *number*
specifies a singularity criterion for the inversion of the matrix. The default depends on the precision of the computer system.

TEST Statement

label: **TEST** *equation [,equation...]*;

The TEST statement performs *F*-tests of linear hypotheses about the regression parameters in the preceding MODEL statement. Each equation specifies a linear hypothesis to be tested. All hypotheses in one TEST statement are tested jointly. Variable names in the equations must correspond to regressors in the preceding MODEL statement, and each name represents the coefficient of the corresponding regressor. The keyword INTERCEPT refers to the coefficient of the intercept.

The following illustrates the use of the TEST statement:

```
proc tscsreg;
   model y = x1 x2 x3;
   test x1 = 0, x2/2 + 2*x3= 0;
   test_int: test intercept=0, x3 = 0;
```

Details

Notation

The discussion here is in the context of the usual panel structure,

$$y_{it} = \sum_{k=1}^{K} x_{itk}\beta_k + u_{it} \quad i = 1, \ldots N; \quad t = 1, \ldots T_i$$

with the specification of u_{it} dependent on the particular model. The total number of observations $M = \sum_{i=1}^{N} T_i$. For the balanced data case, $T_i = T$ for all i. The $M \times M$ covariance matrix of u_{it} is denoted by $\mathbf{V}$. Let $\mathbf{X}$ and $\mathbf{y}$ be the independent and dependent variables arranged by cross section and by time within each cross section. Let $\mathbf{X}_s$ be the X matrix without the intercept. Generally, all other notation is specific to each section.

The One-Way Fixed Effects Model

The specification for the one-way fixed effects model is

$$u_{it} = \nu_i + \epsilon_{it}$$

where the ν_is are nonrandom. Since including both the intercept and all the ν_is induces a redundancy (unless the intercept is suppressed with the NOINT option), the ν_i estimates are reported under the restriction that $\nu_N = 0$.

Let $\mathbf{Q}_0 = diag(\mathbf{E}_{T_i})$, with $\bar{\mathbf{J}}_{T_i} = \mathbf{J}_{T_i}/T_i$ and $\mathbf{E}_{T_i} = \mathbf{I}_{T_i} - \bar{\mathbf{J}}_{T_i}$.

The estimators for the intercept and the fixed effects are given by the usual OLS expressions.

If $\tilde{\mathbf{X}}_s = \mathbf{Q}_0\mathbf{X}_s$ and $\tilde{\mathbf{y}} = \mathbf{Q}_0\mathbf{y}$, the estimator of the slope coefficients is given by

$$\tilde{\beta}_s = (\tilde{\mathbf{X}}'_s\tilde{\mathbf{X}}_s)^{-1}\tilde{\mathbf{X}}'_s\tilde{\mathbf{y}}$$

The estimator of the error variance is

$$\hat{\sigma}_\epsilon = \tilde{\mathbf{u}}'\mathbf{Q}_0\tilde{\mathbf{u}}/(M - N - (K - 1))$$

where the residuals $\tilde{\mathbf{u}}$ are given by $\tilde{\mathbf{u}} = (\mathbf{I}_M - \mathbf{J}_M\mathbf{j}'_M/M)(\mathbf{y} - \mathbf{X}_s\tilde{\beta}_s)$ if there is an intercept and by $\tilde{\mathbf{u}} = (\mathbf{y} - \mathbf{X}_s\tilde{\beta}_s)$if there is not.

The Two-Way Fixed Effects Model

The specification for the two-way fixed effects model is

$$u_{it} = \nu_i + e_t + \epsilon_{it}$$

where the ν_is and e_ts are nonrandom. If you do not specify the NOINT option, which suppresses the intercept, the estimates for the fixed effects are reported under the restriction that $\nu_N = 0$ and $e_T = 0$. If you specify the NOINT option to suppress the intercept, only the restriction $e_T = 0$ is imposed.

Let $\mathbf{X}_*$ and $\mathbf{y}_*$ be the independent and dependent variables arranged by time and by cross section within each time period. Let M_t be the number of cross sections observed in year t and let $\sum_t M_t = M$. Let $\mathbf{D}_t$ be the $M_t \times N$ matrix obtained from the $N \times N$ identity matrix from which rows corresponding to cross sections not observed at time t have been omitted. Consider

$$\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$$

where $\mathbf{Z}_1 = (\mathbf{D}_1', \mathbf{D}_2', \ldots..\mathbf{D}_T')'$ and $\mathbf{Z}_2 = diag(\mathbf{D}_1\mathbf{j}_N, \mathbf{D}_2\mathbf{j}_N, \ldots\ldots\mathbf{D}_T\mathbf{j}_N)$. The matrix $\mathbf{Z}$ gives the dummy variable structure for the two-way model.

Let

$$\Delta_N = \mathbf{Z}_1'\mathbf{Z}_1, \quad \Delta_T = \mathbf{Z}_2'\mathbf{Z}_2, \quad \mathbf{A} = \mathbf{Z}_2'\mathbf{Z}_1$$

$$\bar{\mathbf{Z}} = \mathbf{Z}_2 - \mathbf{Z}_1\Delta_N^{-1}\mathbf{A}'$$

$$\mathbf{Q} = \Delta_T - \mathbf{A}\Delta_N^{-1}\mathbf{A}'$$

$$\mathbf{P} = (\mathbf{I}_M - \mathbf{Z}_1\Delta_N^{-1}\mathbf{Z}_1') - \bar{\mathbf{Z}}\mathbf{Q}^{-}\bar{\mathbf{Z}}'$$

The estimators for the intercept and the fixed effects are given by the usual OLS expressions.

The estimate of the regression slope coefficients is given by

$$\tilde{\beta}_s = (\mathbf{X}_{*s}'\mathbf{P}\mathbf{X}_{*s})^{-1}\mathbf{X}_{*s}'\mathbf{P}\mathbf{y}_*$$

where $\mathbf{X}_{*s}$ is the $\mathbf{X}_*$ matrix without the vector of 1s.

The estimator of the error variance is

$$\hat{\sigma}_\epsilon^2 = \tilde{\mathbf{u}}'\mathbf{P}\tilde{\mathbf{u}}/(M - T - N + 1 - (K - 1))$$

where the residuals are given by $\tilde{\mathbf{u}} = (\mathbf{I}_M - \mathbf{j}_M\mathbf{j}_M'/M)(\mathbf{y}_* - \mathbf{X}_{*s}\tilde{\beta}_s)$if there is an intercept in the model and by $\tilde{\mathbf{u}} = \mathbf{y}_* - \mathbf{X}_{*s}\tilde{\beta}_s$if there is no intercept.

The One-Way Random Effects Model

The specification for the one-way random effects model is

$$u_{it} = \nu_i + \epsilon_{it}$$

Let $\mathbf{Z}_0 = diag(\mathbf{j}_{T_i})$, $\mathbf{P}_0 = diag(\bar{\mathbf{J}}_{T_i})$, and $\mathbf{Q}_0 = diag(\mathbf{E}_{T_i})$, with $\bar{\mathbf{J}}_{T_i} = \mathbf{J}_{T_i}/T_i$ and $\mathbf{E}_{T_i} = \mathbf{I}_{T_i} - \bar{\mathbf{J}}_{T_i}$. Define $\tilde{\mathbf{X}}_s = \mathbf{Q}_0\mathbf{X}_s$ and $\tilde{\mathbf{y}} = \mathbf{Q}_0\mathbf{y}$.

The fixed effects estimator of σ^2_ϵ is still unbiased under the random effects assumptions, so you need to calculate only the estimate of σ_ν.

In the balanced data case, the estimation method for the variance components is the fitting constants method as applied to the one way model; refer to Baltagi and Chang (1994). Fuller and Battese (1974) apply this method to the two-way model.

Let

$$R(\nu) = \mathbf{y}^{'}\mathbf{Z}_0(\mathbf{Z}_0^{'}\mathbf{Z}_0)^{-1}\mathbf{Z}_0^{'}\mathbf{y}$$

$$R(\beta|\nu) = ((\tilde{\mathbf{X}}_s^{'}\tilde{\mathbf{X}}_s)^{-1}\tilde{\mathbf{X}}_s^{'}\tilde{\mathbf{y}})^{'}(\tilde{\mathbf{X}}_s^{'}\tilde{\mathbf{y}})$$

$$R(\beta) = (\mathbf{X}^{'}\mathbf{y})^{'}(\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{X}^{'}\mathbf{y}$$

$$R(\nu|\beta) = R(\beta|\nu) + R(\nu) - R(\beta)$$

The estimator of the error variance is given by

$$\hat{\sigma}^2_\epsilon = (\mathbf{y}^{'}\mathbf{y} - R(\beta|\nu) - R(\nu))/(M - N - (K-1))$$

and the estimator of the cross-sectional variance component is given by

$$\hat{\sigma}^2_\nu = (R(\nu|\beta) - (N-1)\hat{\sigma}^2_\epsilon)/(M - \mathrm{tr}(\mathbf{Z}_0^{'}\mathbf{X}(\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{X}^{'}\mathbf{Z}_0))$$

The estimation of the one-way unbalanced data model is performed using a specialization (Baltagi and Chang 1994) of the approach used by Wansbeek and Kapteyn (1989) for unbalanced two-way models.

The estimation of the variance components is performed by using a quadratic unbiased estimation (QUE) method. This involves focusing on quadratic forms of the centered residuals, equating their expected values to the realized quadratic forms, and solving for the variance components.

Let

$$q_1 = \tilde{\mathbf{u}}^{'}\mathbf{Q}_0\tilde{\mathbf{u}}$$

$$q_2 = \tilde{\mathbf{u}}^{'}\mathbf{P}_0\tilde{\mathbf{u}}$$

where the residuals $\tilde{\mathbf{u}}$ are given by $\tilde{\mathbf{u}} = (\mathbf{I}_M - \mathbf{j}_M\mathbf{j}'_M/M)(\mathbf{y} - \mathbf{X}_s\tilde{\mathbf{X}}'_s\tilde{\mathbf{X}}_s)^{-1}\tilde{\mathbf{X}}_s{}'\tilde{\mathbf{y}})$ if there is an intercept and by $tilde\mathbf{u} = (\mathbf{y} - \mathbf{X}_s(\tilde{\mathbf{X}}'_s\tilde{\mathbf{X}}_s)^{-1}\tilde{\mathbf{X}}'_s\tilde{\mathbf{y}})$ if there is not.

Consider the expected values

$$E(q_1) = (M - N - (K-1))\sigma^2_\epsilon$$

$$E(q_2) = (N - 1 + \mathrm{tr}[(\mathbf{X}_\mathrm{s}'\mathbf{Q}_0\mathbf{X}_\mathrm{s})^{-1}\mathbf{X}_\mathrm{s}'\mathbf{P}_0\mathbf{X}_\mathrm{s}] - \mathrm{tr}[(\mathbf{X}_\mathrm{s}'\mathbf{Q}_0\mathbf{X}_\mathrm{s})^{-1}\mathbf{X}_\mathrm{s}'\bar{\mathbf{J}}_\mathrm{M}\mathbf{X}_\mathrm{s}])\sigma_\epsilon^2$$

$$+[M - (\sum_i T_i^2/M)]\sigma_\nu^2$$

$\hat{\sigma}_\epsilon^2$ and $\hat{\sigma}_\nu^2$ are obtained by equating the quadratic forms to their expected values.

The estimated generalized least squares procedure substitutes the QUE estimates into the covariance matrix of u_{it}, which is given by

$$\mathbf{V} = \sigma_\nu^2 I_M + \sigma_\epsilon^2 \mathbf{Z}_0\mathbf{Z}_0'$$

The Two-Way Random Effects Model

The specification for the two way model is

$$u_{it} = \nu_i + e_t + \epsilon_{it}$$

For balanced data, the two-way random effects model is estimated using the method of Fuller and Battese (1974), so in this case, the RANTWO option is equivalent to the FULLER option already existing in PROC TSCSREG.

The following method (Wansbeek and Kapteyn 1989) is used to handle unbalanced data.

Let $\mathbf{X}_*$ and $\mathbf{y}_*$ be the independent and dependent variables arranged by time and by cross section within each time period. Let M_t be the number of cross sections observed in time t and $\sum_t M_t = M$. Let $\mathbf{D}_t$ be the $M_t \times N$ matrix obtained from the $N \times N$ identity matrix from which rows corresponding to cross sections not observed at time t have been omitted. Consider

$$\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2)$$

where $\mathbf{Z}_1 = (\mathbf{D}_1', \mathbf{D}_2', \ldots..\mathbf{D}_T')'$ and $\mathbf{Z}_2 = diag(\mathbf{D}_1\mathbf{j}_N, \mathbf{D}_2\mathbf{j}_N, \ldots\ldots\mathbf{D}_T\mathbf{j}_N)$.

The matrix $\mathbf{Z}$ gives the dummy variable structure for the two-way model.

Let

$$\Delta_N = \mathbf{Z}_1'\mathbf{Z}_1, \quad \Delta_T = \mathbf{Z}_2'\mathbf{Z}_2, \quad \mathbf{A} = \mathbf{Z}_2'\mathbf{Z}_1$$

$$\bar{\mathbf{Z}} = \mathbf{Z}_2 - \mathbf{Z}_1\Delta_N^{-1}\mathbf{A}'$$

$$\mathbf{Q} = \Delta_T - \mathbf{A}\Delta_N^{-1}\mathbf{A}'$$

$$\mathbf{P} = (\mathbf{I}_M - \mathbf{Z}_1\Delta_N^{-1}\mathbf{Z}_1') - \bar{\mathbf{Z}}\mathbf{Q} - \bar{\mathbf{Z}}'$$

The estimator of the error variance is

$$\hat{\sigma}_{\epsilon}^{2} = \tilde{\mathbf{u}}'\mathbf{P}\tilde{\mathbf{u}}/M - T - N + 1 - (K - 1))$$

where the $\tilde{\mathbf{u}}$ are given by $\tilde{\mathbf{u}} = (\mathbf{I}_M - \mathbf{j}_M\mathbf{j}'_M/M)(\mathbf{y}_* - \mathbf{X}_{*s}(\mathbf{X}'_{*s}\mathbf{P}\mathbf{X}_{*s})^{-1}\mathbf{X}_{*s}'\mathbf{P}\mathbf{y}_*)$ if there is an intercept and by $\tilde{\mathbf{u}} = (\mathbf{y}_* - \mathbf{X}_{*s}(\mathbf{X}'_{*s}\mathbf{P}\mathbf{X}_{*s})^{-1}\mathbf{X}'_{*s}\mathbf{P}\mathbf{y}_*$if there is not.

The estimation of the variance components is performed by using a quadratic unbiased estimation (QUE) method that involves focusing on quadratic forms of the residuals $\tilde{\mathbf{u}}$, equating their expected values to the realized quadratic forms, and solving for the variance components.

Let

$$q_N = \tilde{\mathbf{u}}'\mathbf{Z_2}\Delta_T^{-1}\mathbf{Z}_2'\tilde{\mathbf{u}}$$

$$q_T = \tilde{\mathbf{u}}'\mathbf{Z}_1\Delta_N^{-1}\mathbf{Z}_1'\tilde{\mathbf{u}}$$

Consider the expected values

$$E(q_N) = (T + k_N - (1 + k_0))\sigma^2 + (T - \frac{\lambda_1}{M})\sigma_{\nu}^2 + (M - \frac{\lambda_2}{M})\sigma_e^2$$

$$E(q_T) = (N + k_T - (1 + k_0))\sigma^2 + (M - \frac{\lambda_1}{M})\sigma_{\nu}^2 + (N - \frac{\lambda_2}{M})\sigma_e^2$$

where

$$k_0 = \mathbf{j}_M'\mathbf{X}_{*s}(\mathbf{X}_{*s}'\mathbf{P}\mathbf{X}_{*s})^{-1}\mathbf{X}_{*s}'\mathbf{j}_M/M$$

$$k_N = tr((\mathbf{X}_{*s}'\mathbf{P}\mathbf{X}_{*s})^{-1}\mathbf{X}_{*s}'\mathbf{Z}_2\Delta_T^{-1}\mathbf{Z}_2'\mathbf{X}_{*s})$$

$$k_T = tr((\mathbf{X}_{*s}'\mathbf{P}\mathbf{X}_{*s})^{-1}\mathbf{X}_{*s}'\mathbf{Z}_1\Delta_N^{-1}\mathbf{Z}_1'\mathbf{X}_{*s})$$

$$\lambda_1 = \mathbf{j}_M'\mathbf{Z}_1\mathbf{Z}_1'\mathbf{j}_M$$

$$\lambda_2 = \mathbf{j}_M'\mathbf{Z}_2\mathbf{Z}_2'\mathbf{j}_M$$

The quadratic unbiased estimators for σ_{ν}^2 and σ_e^2 are obtained by equating the expected values to the quadratic forms and solving for the two unknowns.

The estimated generalized least squares procedure substitute the QUE estimates into the covariance matrix of the composite error term u_{it}, which is given by

$$\mathbf{V} = \sigma_{\epsilon}^2\mathbf{I}_M + \sigma_{\nu}^2\mathbf{Z}_1\mathbf{Z}_1' + \sigma_e^2\mathbf{Z}_2\mathbf{Z}_2'$$

Parks Method (Autoregressive Model)

Parks (1967) considered the first-order autoregressive model in which the random errors u_{it}, $\quad i = 1, 2, \ldots, N, \quad t = 1, 2, \ldots, T$, have the structure

$$\begin{aligned} E(u_{it}^2) &= \sigma_{ii} && \text{(heteroscedasticity)} \\ E(u_{it}u_{jt}) &= \sigma_{ij} && \text{(contemporaneously correlated)} \\ u_{it} &= \rho_i u_{i,t-1} + \epsilon_{it} && \text{(autoregression)} \end{aligned}$$

where

$$\begin{aligned} E(\epsilon_{it}) &= 0 \\ E(u_{i,t-1}\epsilon_{jt}) &= 0 \\ E(\epsilon_{it}\epsilon_{jt}) &= \phi_{ij} \\ E(\epsilon_{it}\epsilon_{js}) &= 0 \qquad (s \neq t) \\ E(u_{i0}) &= 0 \\ E(u_{i0}u_{j0}) &= \sigma_{ij} = \phi_{ij}/(1 - \rho_i\rho_j) \end{aligned}$$

The model assumed is first-order autoregressive with contemporaneous correlation between cross sections. In this model, the covariance matrix for the vector of random errors **u** can be expressed as

$$E(\mathbf{u}\mathbf{u}^{'}) = \mathbf{V} = \begin{bmatrix} \sigma_{11}P_{11} & \sigma_{12}P_{12} & \ldots & \sigma_{1N}P_{1N} \\ \sigma_{21}P_{21} & \sigma_{22}P_{22} & \ldots & \sigma_{2N}P_{2N} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{N1}P_{N1} & \sigma_{N2}P_{N2} & \ldots & \sigma_{NN}P_{NN} \end{bmatrix}$$

where

$$P_{ij} = \begin{bmatrix} 1 & \rho_j & \rho_j^2 & \ldots & \rho_j^{T-1} \\ \rho_i & 1 & \rho_j & \ldots & \rho_j^{T-2} \\ \rho_i^2 & \rho_i & 1 & \ldots & \rho_j^{T-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho_i^{T-1} & \rho_i^{T-2} & \rho_i^{T-3} & \ldots & 1 \end{bmatrix}$$

The matrix **V** is estimated by a two-stage procedure, and β is then estimated by generalized least squares. The first step in estimating **V** involves the use of ordinary least squares to estimate β and obtain the fitted residuals, as follows:

$$\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}\hat{\beta}_{OLS}$$

A consistent estimator of the first-order autoregressive parameter is then obtained in the usual manner, as follows:

$$\hat{\rho}_i = \left(\sum_{t=2}^{T} \hat{u}_{it} \hat{u}_{i,t-1} \right) \Bigg/ \left(\sum_{t=2}^{T} \hat{u}_{i,t-1}^2 \right) \quad i = 1, 2, \ldots, N$$

Finally, the autoregressive characteristic of the data can be removed (asymptotically) by the usual transformation of taking weighted differences. That is, for $i = 1, 2, \ldots, N$,

$$y_{i1} \sqrt{1 - \hat{\rho}_i^2} = \sum_{k=1}^{p} X_{i1k} \beta_k \sqrt{1 - \hat{\rho}_i^2} + u_{i1} \sqrt{1 - \hat{\rho}_i^2}$$

$$y_{it} - \hat{\rho}_i y_{i,t-1} = \sum_{k=1}^{p} (X_{itk} - \hat{\rho}_i \mathbf{X}_{i,t-1,k}) \beta_k + u_{it} - \hat{\rho}_i u_{i,t-1} \quad t = 2, \ldots, T$$

which is written

$$y_{it}^* = \sum_{k=1}^{p} X_{itk}^* \beta_k + u_{it}^* \quad i = 1, 2, \ldots, N; \quad t = 1, 2, \ldots, T$$

Notice that the transformed model has not lost any observations (Seely and Zyskind 1971).

The second step in estimating the covariance matrix **V** is to apply ordinary least squares to the preceding transformed model, obtaining

$$\hat{\mathbf{u}}^* = \mathbf{y}^* - \mathbf{X}^* \beta_{OLS}^*$$

from which the consistent estimator of σ_{ij} is calculated:

$$s_{ij} = \frac{\hat{\phi}_{ij}}{(1 - \hat{\rho}_i \hat{\rho}_j)}$$

where

$$\hat{\phi}_{ij} = \frac{1}{(T - p)} \sum_{t=1}^{T} \hat{u}_{it}^* \hat{u}_{jt}^*$$

EGLS then proceeds in the usual manner,

$$\hat{\beta}_P = (\mathbf{X}' \hat{\mathbf{V}}^{-1} \mathbf{X})^{-1} \mathbf{X}' \hat{\mathbf{V}}^{-1} \mathbf{y}$$

where $\hat{\mathbf{V}}$ is the derived consistent estimator of $\mathbf{V}$. For computational purposes, it should be pointed out that $\hat{\beta}_P$ is obtained directly from the transformed model,

$$\hat{\beta}_P = (\mathbf{X}^{*'}(\hat{\Phi}^{-1}\otimes I_T)\mathbf{X}^{*})^{-1}\mathbf{X}^{*'}(\hat{\Phi}^{-1}\otimes I_T)\mathbf{y}^{*}$$

where $\hat{\Phi} = [\hat{\phi}_{ij}]_{i,j=1,\ldots,N}$.

The preceding procedure is equivalent to Zellner's two-stage methodology applied to the transformed model (Zellner 1962).

Parks demonstrates that his estimator is consistent and asymptotically, normally distributed with

$$\mathrm{Var}(\hat{\beta}_{\mathrm{P}}) = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}$$

Standard Corrections

For the PARKS option, the first-order autocorrelation coefficient must be estimated for each cross section. Let ρ be the $N*1$ vector of true parameters and $R = (r_1, \ldots, r_N)'$ be the corresponding vector of estimates. Then, to ensure that only range-preserving estimates are used in PROC TSCSREG, the following modification for R is made:

$$r_i = \begin{cases} r_i & \text{if } |r_i| < 1 \\ max(.95, rmax) & \text{if } r_i \geq 1 \\ min(-.95, rmin) & \text{if } r_i \leq -1 \end{cases}$$

where

$$rmax = \begin{cases} 0 & \text{if } r_i < 0 \text{ or } r_i \geq 1 \text{ for all i} \\ \max_j [r_j : 0 \leq r_j < 1] & \text{otherwise} \end{cases}$$

and

$$rmin = \begin{cases} 0 & \text{if } r_i > 0 \text{ or } r_i \leq -1 \text{ for all i} \\ \max_j [r_j : -1 < r_j \leq 0] & \text{otherwise} \end{cases}$$

Whenever this correction is made, a warning message is printed.

Da Silva Method (Variance-Component Moving Average Model)

Suppose you have a sample of observations at *T* time points on each of *N* cross-sectional units. The Da Silva method assumes that the observed value of the dependent variable at the *t*th time point on the *i*th cross-sectional unit can be expressed as

$$y_{it} = \mathbf{x}_{it}'\beta + a_i + b_t + e_{it} \quad i = 1, \ldots, N; \quad t = 1, \ldots, T$$

where

$\mathbf{x}_{it}^{'} = (x_{it1}, \ldots, x_{itp})$is a vector of explanatory variables for the *t*th time point and *i*th cross-sectional unit

$\beta = (\beta_1, \ldots, \beta_p)'$ is the vector of parameters

a_i is a time-invariant, cross-sectional unit effect

b_t is a cross-sectionally invariant time effect

e_{it} is a residual effect unaccounted for by the explanatory variables and the specific time and cross-sectional unit effects

Since the observations are arranged first by cross sections, then by time periods within cross sections, these equations can be written in matrix notation as

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u}$$

where

$$\mathbf{u} = (\mathbf{a} \otimes \mathbf{1}_T) + (\mathbf{1}_N \otimes \mathbf{b}) + \mathbf{e}$$

$$\mathbf{y} = (y_{11}, \ldots, y_{1T}, y_{21}, \ldots, y_{NT})'$$

$$\mathbf{X} = (\mathbf{x}_{11}, \ldots, \mathbf{x}_{1T}, \mathbf{x}_{21}, \ldots, \mathbf{x}_{NT})'$$

$$\mathbf{a} = (a_1 \ldots a_N)'$$

$$\mathbf{b} = (b_1 \ldots b_T)'$$

$$\mathbf{e} = (e_{11}, \ldots, e_{1T}, e_{21}, \ldots, e_{NT})'$$

Here $\mathbf{1}_N$ is an $N \times 1$ vector with all elements equal to 1, and $\otimes$ denotes the Kronecker product.

It is assumed that

1. $\mathbf{x}_{it}$ is a sequence of nonstochastic, known $p \times 1$ vectors in $\Re^p$ whose elements are uniformly bounded in $\Re^p$. The matrix **X** has a full column rank *p*.
2. β is a $p \times 1$ constant vector of unknown parameters.
3. **a** is a vector of uncorrelated random variables such that $E(a_i) = 0$ and $var(a_i) = \sigma_a^2, \sigma_a^2 > 0, i = 1, \ldots, N$.
4. **b** is a vector of uncorrelated random variables such that $E(b_t) = 0$ and $var(b_t) = \sigma_b^2, \sigma_b^2 > 0, t = 1, \ldots, T$.

5. $\mathbf{e}_i = (e_{i1}, \ldots, e_{iT})'$ is a sample of a realization of a finite moving average time series of order $m < T - 1$ for each *i*; hence,

$$e_{it} = \alpha_0 \epsilon_t + \alpha_1 \epsilon_{t-1} + \ldots + \alpha_m \epsilon_{t-m}, \quad t = 1, \ldots, T; \quad i = 1, \ldots, N$$

   where $\alpha_0, \alpha_1, \ldots, \alpha_m$ are unknown constants such that $\alpha_0 \neq 0$ and $\alpha_m \neq 0$, and $\{\epsilon_j\}_{j=-\infty}^{j=\infty}$ is a white noise process, that is, a sequence of uncorrelated random variables with $E(\epsilon_t) = 0, E(\epsilon_t^2) = \sigma_\epsilon^2$, and $\sigma_\epsilon^2 > 0$.

6. The sets of random variables $\{a_i\}_{i=1}^N$, $\{b_t\}_{t=1}^T$, and $\{e_{it}\}_{t=1}^T$ for $i = 1, \ldots, N$ are mutually uncorrelated.

7. The random terms have normal distributions: $a_i \sim N(0, \sigma_a^2), b_t \sim N(0, \sigma_b^2)$,and $\epsilon_{t-k} \sim N(0, \sigma_\epsilon^2)$, for $i = 1, \ldots, N; t = 1, \ldots T; k = 1, \ldots, m$.

If assumptions 1-6 are satisfied, then

$$E(\mathbf{y}) = \mathbf{X}\beta$$

and

$$var(\mathbf{y}) = \sigma_a^2 (I_N \otimes J_T) + \sigma_b^2 (J_N \otimes I_T) + (I_N \otimes \Gamma_T)$$

where Γ_T is a $T \times T$ matrix with elements γ_{ts} as follows:

$$cov(e_{it} e_{is}) = \begin{cases} \gamma(|t-s|) & \text{if } |t-s| \leq \text{m} \\ 0 & \text{if } |t-s| > \text{m} \end{cases}$$

where $\gamma(k) = \sigma_\epsilon^2 \sum_{j=0}^{m-k} \alpha_j \alpha_{j+k}$ for $k = |t-s|$. For the definition of I_N, I_T, J_N, and J_T, see the "Fuller-Battese Method" section earlier in this chapter.

The covariance matrix, denoted by **V**, can be written in the form

$$\mathbf{V} = \sigma_a^2 (I_N \otimes J_T) + \sigma_b^2 (J_N \otimes I_T) + \sum_{k=0}^{m} \gamma(k) (I_N \otimes \Gamma_T^{(k)})$$

where $\Gamma_T^{(0)} = I_T$, and, for *k*=1,..., *m*, $\Gamma_T^{(k)}$ is a band matrix whose *k*th off-diagonal elements are 1's and all other elements are 0's.

Thus, the covariance matrix of the vector of observations **y** has the form

$$var(\mathbf{y}) = \sum_{k=1}^{m+3} \nu_k V_k$$

where

$$\begin{aligned}
\nu_1 &= \sigma_a^2 \\
\nu_2 &= \sigma_b^2 \\
\nu_k &= \gamma(k-3) \qquad k = 3, \ldots, m+3 \\
V_1 &= I_N \otimes J_T \\
V_2 &= J_N \otimes I_T \\
V_k &= I_N \otimes \Gamma_T^{(k-3)} \quad k = 3, \ldots, m+3
\end{aligned}$$

The estimator of β is a two-step GLS-type estimator, that is, GLS with the unknown covariance matrix replaced by a suitable estimator of **V**. It is obtained by substituting Seely estimates for the scalar multiples $\nu_k, k = 1, 2, \ldots, m+3$.

Seely (1969) presents a general theory of unbiased estimation when the choice of estimators is restricted to finite dimensional vector spaces, with a special emphasis on quadratic estimation of functions of the form $\sum_{i=1}^{n} \delta_i \nu_i$.

The parameters ν_i (i=1,. . ., n) are associated with a linear model E(**y**)=**X**β with covariance matrix $\sum_{i=1}^{n} \nu_i V_i$ where V_i (i=1, . . ., n) are real symmetric matrices. The method is also discussed by Seely (1970a,1970b) and Seely and Zyskind (1971). Seely and Soong (1971) consider the MINQUE principle, using an approach along the lines of Seely (1969).

Linear Hypothesis Testing

For a linear hypothesis of the form **R** β=**r** where **R** is $J{\times}L$ and **r** is $J{\times}1$, the F-statistic with $J, M-L$ degrees of freedom is computed as

$$(\mathbf{R}\beta - \mathbf{r})'[\mathbf{R}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{R}']^{-1}\mathbf{R}(\mathbf{R}\beta - \mathbf{r})$$

R-squared

The conventional R-squared measure is inappropriate for all models that the TSCSREG procedure estimates using GLS since a number outside the 0-to-1 range may be produced. Hence, a generalization of the R-squared measure is reported. The following goodness-of-fit measure (Buse 1973) is reported:

$$R^2 = 1 - \frac{\hat{\mathbf{u}}'\hat{\mathbf{V}}^{-1}\hat{\mathbf{u}}}{\mathbf{y}'\mathbf{D}'\hat{\mathbf{V}}^{-1}\mathbf{D}\mathbf{y}}$$

where $\hat{\mathbf{u}}$ are the residuals of the transformed model, $\hat{\mathbf{u}} = \mathbf{y} - \mathbf{X}(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{y}$,

and $\mathbf{D} = \mathbf{I}_M - \mathbf{j}_M\mathbf{j}_M'(\frac{\hat{\mathbf{V}}^{-1}}{\mathbf{j}_M'\hat{\mathbf{V}}^{-1}\mathbf{j}_M})$.

This is a measure of the proportion of the transformed sum of squares of the dependent variable that is attributable to the influence of the independent variables.

If there is no intercept in the model, the corresponding measure (Theil 1961) is

$$R^2 = 1 - \frac{\hat{\mathbf{u}}'\hat{\mathbf{V}}^{-1}\hat{\mathbf{u}}}{\mathbf{y}'\hat{\mathbf{V}}^{-1}\mathbf{y}}$$

Clearly, in the case of OLS estimation, both the R-squared formulas given here reduce to the usual R-squared formula.

Specification Tests

The TSCSREG procedure outputs the results of one specification test for fixed effects and one specification test for random effects.

For fixed effects, let β_f be the n dimensional vector of fixed effects parameters. The specification test reported is the conventional *F*-statistic for the hypothesis $\beta_f = \mathbf{0}$. The *F*-statistic with $n, M - K$ degrees of freedom is computed as

$$\hat{\beta}_f \hat{\mathbf{S}}_f^{-1} \hat{\beta}_f / n$$

where $\hat{\mathbf{S}}_f$ is the estimated covariance matrix of the fixed effects parameters.

Hausman's (1978) specification test or *m*-statistic can be used to test hypotheses in terms of bias or inconsistency of an estimator. This test was also proposed by Wu (1973) and further extended in Hausman and Taylor (1982). Hausman's *m*-statistic is as follows.

Consider two estimators, $\hat{\beta}_a$ and $\hat{\beta}_b$, which under the null hypothesis are both consistent, but only $\hat{\beta}_a$ is asymptotically efficient. Under the alternative hypothesis, only $\hat{\beta}_b$ is consistent. The *m*-statistic is

$$m = (\hat{\beta}_b - \hat{\beta}_a)'(\hat{\mathbf{S}}_b - \hat{\mathbf{S}}_a)^{-}(\hat{\beta}_b - \hat{\beta}_a)$$

where $\hat{\mathbf{S}}_b$ and $\hat{\mathbf{S}}_a$ are consistent estimates of the asymptotic covariance matrices of $\hat{\beta}_b$ and $\hat{\beta}_a$. Then m is distributed χ^2 with k degrees of freedom, where k is the dimension of $\hat{\beta}_a$ and $\hat{\beta}_b$.

In the random effects specification, the null hypothesis of no correlation between effects and regressors implies that the OLS estimates of the slope parameters are consistent and inefficient but the GLS estimates of the slope parameters are consistent and efficient. This facilitates a Hausman specification test. The reported χ^2 statistic has degrees of freedom equal to the number of slope parameters.

OUTEST= Data Set

PROC TSCSREG writes the parameter estimates to an output data set when the OUTEST= option is specified. The OUTEST= data set contains the following variables:

MODEL a character variable containing the label for the MODEL statement if a label is specified

METHOD a character variable identifying the estimation method. Current methods are FULLER, PARKS, and DASILVA.

TYPE a character variable that identifies the type of observation. Values of the _TYPE_ variable are CORRB, COVB, CSPARMS, and PARMS; the CORRB observation contains correlations of the parameter estimates; the COVB observation contains covariances of the parameter estimates; the CSPARMS observation contains cross-sectional parameter estimates; and the PARMS observation contains parameter estimates.

NAME a character variable containing the name of a regressor variable for COVB and CORRB observations and left blank for other observations. The _NAME_ variable is used in conjunction with the _TYPE_ values COVB and CORRB to identify rows of the correlation or covariance matrix.

DEPVAR a character variable containing the name of the response variable

MSE the mean square error of the transformed model

CSID the value of the cross section ID for CSPARMS observations. _CSID_ is used with the _TYPE_ value CSPARMS to identify the cross section for the first order autoregressive parameter estimate contained in the observation. _CSID_ is missing for observations with other _TYPE_ values. (Currently only the _A_1 variable contains values for CSPARMS observations.)

VARCS the variance component estimate due to cross sections. _VARCS_ is included in the OUTEST= data set when either the FULLER or DASILVA option is specified.

VARTS the variance component estimate due to time series. _VARTS_ is included in the OUTEST= data set when either the FULLER or DASILVA option is specified.

VARERR the variance component estimate due to error. _VARERR_ is included in the OUTEST= data set when the FULLER option is specified.

_A_1 the first order autoregressive parameter estimate. _A_1 is included in the OUTEST= data set when the PARKS option is specified. The values of _A_1 are cross-sectional parameters, meaning that they are estimated for each cross section separately. _A_1 has a value only for _TYPE_=CSPARMS observations. The cross section to which the estimate belongs is indicated by the _CSID_ variable.

INTERCEP the intercept parameter estimate. (INTERCEP will be missing for models for which the NOINT option is specified.)

regressors — the regressor variables specified in the MODEL statement. The regressor variables in the OUTEST= data set contain the corresponding parameter estimates for the model identified by _MODEL_ for _TYPE_=PARMS observations, and the corresponding covariance or correlation matrix elements for _TYPE_=COVB and _TYPE_=CORRB observations. The response variable contains the value -1 for the _TYPE_=PARMS observation for its model.

Printed Output

For each MODEL statement, the printed output from PROC TSCSREG includes the following:

1. a model description, which gives the estimation method used, the model statement label if specified, the number of cross sections and the number of observations in each cross section, and the order of moving average error process for the DASILVA option
2. the estimates of the underlying error structure parameters
3. the regression parameter estimates and analysis. For each regressor, this includes the name of the regressor, the degrees of freedom, the parameter estimate, the standard error of the estimate, a *t* statistic for testing whether the estimate is significantly different from 0, and the significance probability of the *t* statistic. Whenever possible, the notation of the original reference is followed.

Optionally, PROC TSCSREG prints the following:

4. the covariance and correlation of the resulting regression parameter estimates for each model and assumed error structure
5. the $\hat{\Phi}$ matrix that is the estimated contemporaneous covariance matrix for the PARKS option

ODS Table Names

PROC TSCSREG 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 20.1. ODS Tables Produced in PROC TSCSREG

ODS Table Name	Description	Option
ODS Tables Created by the MODEL Statement		
ModelDescription	Model Description	
FitStatistics	Fit Statistics	
FixedEffectsTest	F Test for No Fixed Effects	
ParameterEstimates	Parameter Estimates	
CovB	Covariance of Parameter Estimates	
CorrB	Correlations of Parameter Estimates	
VarianceComponents	Variance Component Estimates	
RandomEffectsTest	Hausman Test for Random Effects	
AR1Estimates	First Order Autoregressive Parameter Estimates	
EstimatedPhiMatrix	Estimated Phi Matrix	PARKS
EstimatedAutocovariances	Estimates of Autocovariances	PARKS
ODS Tables Created by the TEST Statement		
TestResults	Test Results	

Example

Example 20.1. Analyzing Demand for Liquid Assets

In this example, the demand equations for liquid assets are estimated. The demand function for the demand deposits is estimated under three error structures while demand equations for time deposits and savings and loan (S & L) association shares are calculated using the Parks method. The data for seven states (CA, DC, FL, IL, NY, TX, and WA) are selected out of 49 states. Refer to Feige (1964) for data description. All variables were transformed via natural logarithm. The first five observations of the data set A are shown in Output 20.1.1.

```
data a;
   input state $ year d t s y rd rt rs;
   label d = 'Per Capita Demand Deposits'
         t = 'Per Capita Time Deposits'
         s = 'Per Capita S & L Association Shares'
         y = 'Permanent Per Capita Personal Income'
         rd = 'Service Charge on Demand Deposits'
         rt = 'Interest on Time Deposits'
         rs = 'Interest on S & L Association Shares';
datalines;
   ... data lines are omitted ...
;

proc print data=a(obs=5);
run;
```

Output 20.1.1. A Sample of Liquid Assets Data

Obs	state	year	d	t	s	y	rd	rt	rs
1	CA	1949	6.2785	6.1924	4.4998	7.2056	-1.0700	0.1080	1.0664
2	CA	1950	6.4019	6.2106	4.6821	7.2889	-1.0106	0.1501	1.0767
3	CA	1951	6.5058	6.2729	4.8598	7.3827	-1.0024	0.4008	1.1291
4	CA	1952	6.4785	6.2729	5.0039	7.4000	-0.9970	0.4492	1.1227
5	CA	1953	6.4118	6.2538	5.1761	7.4200	-0.8916	0.4662	1.2110

The SORT procedure is used to sort the data into the required time series cross-sectional format. Then PROC TSCSREG analyzes the data.

```
proc sort data=a;
   by state year;
run;

title 'Demand for Liquid Assets';
proc tscsreg data=a;
   model d = y rd rt rs / fuller parks dasilva m=7;
   model t = y rd rt rs / parks;
   model s = y rd rt rs / parks;
   id state year;
run;
```

The income elasticities for liquid assets are greater than 1 except for the demand deposit income elasticity (0.692757) estimated by the Da Silva method. In Output 20.1.2, Output 20.1.3 and Output 20.1.4, the coefficient estimates (-0.29094, -0.43591, and -0.27736) of demand deposits (RD) imply that demand deposits increase significantly as the service charge is reduced. The price elasticities (0.227152 and 0.408066) for time deposits (RT) and S & L association shares (RS) have the expected sign and thus an increase in the interest rate on time deposits or S & L shares will increase the demand for the corresponding liquid asset. Demand deposits and S & L shares appear to be substitutes (Output 20.1.2, Output 20.1.3, Output 20.1.4, and Output 20.1.6). Time deposits are also substitutes for S & L shares in the time deposit demand equation (Output 20.1.5), while these liquid assets are independent of each other in Output 20.1.6 (insignificant coefficient estimate of RT, -0.02705). Demand deposits and time deposits appear to be weak complements in Output 20.1.3 and Output 20.1.4, while the cross elasticities between demand deposits and time deposits are not significant in Output 20.1.2 and Output 20.1.5.

Output 20.1.2. Demand for Demand Deposits – Fuller-Battese Method

```
                        Demand for Liquid Assets

                         The TSCSREG Procedure
                  Fuller and Battese Method Estimation

Dependent Variable: d Per Capita Demand Deposits

                           Model Description

                  Estimation Method              Fuller
                  Number of Cross Sections            7
                  Time Series Length                 11


                             Fit Statistics

               SSE              0.0795    DFE                 72
               MSE              0.0011    Root MSE        0.0332
               R-Square         0.6786


                      Variance Component Estimates

               Variance Component for Cross Sections     0.03427
               Variance Component for Time Series        0.00026
               Variance Component for Error              0.00111


                             Hausman Test for
                              Random Effects

                           DF    m Value    Pr > m

                            4       5.51    0.2385


                          Parameter Estimates

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

 Intercept      1  -1.23606    0.7252    -1.70    0.0926  Intercept
 y              1  1.064058    0.1040    10.23    <.0001  Permanent Per Capita
                                                          Personal Income
 rd             1  -0.29094    0.0526    -5.53    <.0001  Service Charge on
                                                          Demand Deposits
 rt             1  0.039388    0.0278     1.42    0.1603  Interest on Time
                                                          Deposits
 rs             1  -0.32662    0.1140    -2.86    0.0055  Interest on S & L
                                                          Association Shares
```

Output 20.1.3. Demand for Demand Deposits – Parks Method

```
                            Demand for Liquid Assets

                             The TSCSREG Procedure
                            Parks Method Estimation

Dependent Variable: d Per Capita Demand Deposits

                               Model Description

                      Estimation Method             Parks
                      Number of Cross Sections          7
                      Time Series Length               11


                                Fit Statistics

               SSE             73.3696    DFE                   72
               MSE              1.0190    Root MSE          1.0095
               R-Square         0.9263


                              Parameter Estimates

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

 Intercept      1  -2.66565         0.3139    -8.49    <.0001  Intercept
 y              1  1.222569         0.0423    28.87    <.0001  Permanent Per Capita
                                                               Personal Income
 rd             1  -0.43591         0.0201   -21.71    <.0001  Service Charge on
                                                               Demand Deposits
 rt             1  0.041237         0.0210     1.97    0.0530  Interest on Time
                                                               Deposits
 rs             1  -0.26683         0.0654    -4.08    0.0001  Interest on S & L
                                                               Association Shares
```

Output 20.1.4. Demand for Demand Deposits – Da Silva Method

```
                          Demand for Liquid Assets

                           The TSCSREG Procedure
                         Da Silva Method Estimation

Dependent Variable: d Per Capita Demand Deposits

                             Model Description

                     Estimation Method            DaSilva
                     Number of Cross Sections           7
                     Time Series Length                11
                     Order of MA Error Process          7


                              Fit Statistics

               SSE          21609.8923    DFE                  72
               MSE            300.1374    Root MSE        17.3245
               R-Square         0.4995


                        Variance Component Estimates

               Variance Component for Cross Sections     0.03063
               Variance Component for Time Series       0.000148


                              Estimates of
                             Autocovariances
                            Lag            Gamma

                            0        0.0008558553
                            1        0.0009081747
                            2        0.0008494797
                            3        0.0007889687
                            4        0.0013281983
                            5        0.0011091685
                            6        0.0009874973
                            7        0.0008462601
```

```
                          Demand for Liquid Assets

                           The TSCSREG Procedure
                         Da Silva Method Estimation

Dependent Variable: d Per Capita Demand Deposits

                            Parameter Estimates

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

 Intercept      1  1.281084    0.0824    15.55    <.0001  Intercept
 y              1  0.692757   0.00677   102.40    <.0001  Permanent Per Capita
                                                          Personal Income
 rd             1  -0.27736   0.00274  -101.18    <.0001  Service Charge on
                                                          Demand Deposits
 rt             1  0.009378   0.00171     5.49    <.0001  Interest on Time
                                                          Deposits
 rs             1  -0.09942   0.00601   -16.53    <.0001  Interest on S & L
                                                          Association Shares
```

Output 20.1.5. Demand for Time Deposits – Parks Method

```
                            Demand for Liquid Assets

                             The TSCSREG Procedure
                            Parks Method Estimation

Dependent Variable: t Per Capita Time Deposits

                               Model Description

                      Estimation Method              Parks
                      Number of Cross Sections           7
                      Time Series Length                11


                                 Fit Statistics

               SSE               63.3807    DFE                   72
               MSE                0.8803    Root MSE          0.9382
               R-Square           0.9517


                              Parameter Estimates

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

 Intercept      1  -5.33334    0.5007   -10.65    <.0001  Intercept
 y              1  1.516344    0.0810    18.72    <.0001  Permanent Per Capita
                                                          Personal Income
 rd             1  -0.04791    0.0294    -1.63    0.1082  Service Charge on
                                                          Demand Deposits
 rt             1  0.227152    0.0332     6.85    <.0001  Interest on Time
                                                          Deposits
 rs             1  -0.42569    0.1262    -3.37    0.0012  Interest on S & L
                                                          Association Shares
```

Output 20.1.6. Demand for Savings and Loan Shares – Parks Method

```
                           Demand for Liquid Assets

                            The TSCSREG Procedure
                           Parks Method Estimation

Dependent Variable: s Per Capita S & L Association Shares

                              Model Description

                      Estimation Method              Parks
                      Number of Cross Sections           7
                      Time Series Length                11


                               Fit Statistics

               SSE             71.9675    DFE                   72
               MSE              0.9995    Root MSE          0.9998
               R-Square         0.9017


                             Parameter Estimates

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

 Intercept      1  -8.09632      0.7850   -10.31    <.0001  Intercept
 y              1  1.832988      0.1157    15.84    <.0001  Permanent Per Capita
                                                            Personal Income
 rd             1  0.576723      0.0435    13.26    <.0001  Service Charge on
                                                            Demand Deposits
 rt             1  -0.02705      0.0312    -0.87    0.3891  Interest on Time
                                                            Deposits
 rs             1  0.408066      0.1092     3.74    0.0004  Interest on S & L
                                                            Association Shares
```

Acknowledgements

The TSCSREG procedure was developed by Douglas J. Drummond and A. Ronald Gallant, and contributed to the Version 5 SUGI Supplemental Library in 1979.

Dr. Drummond, now deceased, was with the Center for Survey Statistics, Research Triangle Park, NC. Dr. Drummond programmed the Parks and Fuller-Battese methods. Professor Gallant, who is currently with the University of North Carolina at Chapell Hill, programmed the Da Silva method and generously contributed his time to the support of PROC TSCSREG after Dr. Drummond's death.

The version of PROC TSCSREG documented here was produced by converting the older SUGI Supplemental Library version of the procedure to Version 6 of SAS software. This conversion work was performed by SAS Institute, which now supports the procedure. Although several features were added during the conversion (such as the OUTEST= option, ID statement, and BY statement), credit for the statistical aspects and general design of the TSCSREG procedure belongs to Dr. Drummond and Professor Gallant.

References

Baltagi, B. H. and Chang, Y. (1994), "Incomplete Panels: A Comparative Study of Alternative Estimators for the Unbalanced One-way Error Component Regression Model," *Journal of Econometrics*, 62(2), 67-89.

Buse, A. (1973), "Goodness of Fit in Generalized Least Squares Estimation," *American Statistician*, 27, 106-108.

Da Silva, J.G.C. (1975), "The Analysis of Cross-Sectional Time Series Data," Ph.D. dissertation, Department of Statistics, North Carolina State University.

SAS Institute Inc. (1979), *SAS Technical Report S-106, TSCSREG: A SAS Procedure for the Analysis of Time-Series Cross-Section Data*, Cary, NC: SAS Institute Inc.

Feige, E.L. (1964), *The Demand for Liquid Assets: A Temporal Cross-Section Analysis*, Englewood Cliffs: Prentice-Hall.

Feige, E.L. and Swamy, P.A.V. (1974), "A Random Coefficient Model of the Demand for Liquid Assets," *Journal of Money, Credit, and Banking*, 6, 241-252.

Fuller, W.A. and Battese, G.E. (1974), "Estimation of Linear Models with Crossed-Error Structure," *Journal of Econometrics*, 2, 67-78.

Greene, W.H. (1990), *Econometric Analysis*, New York: Macmillan Publishing Company.

Hausman, J.A. (1978), "Specification Tests in Econometrics," *Econometrica*, 46, 1251-1271.

Hausman, J.A. and Taylor, W.E. (1982), "A Generalized Specification Test," *Economics Letters*, 8, 239-245.

Hsiao, C. (1986), *Analysis of Panel Data*, Cambridge: Cambridge University Press.

Judge, G.G., Griffiths, W.E., Hill, R.C., Lutkepohl, H., and Lee, T.C. (1985), *The Theory and Practice of Econometrics*, Second Edition, New York: John Wiley & Sons.

Kmenta, J. (1971), *Elements of Econometrics*, New York: MacMillan Publishing Company, Inc.

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

Parks, R.W. (1967), "Efficient Estimation of a System of Regression Equations when Disturbances Are Both Serially and Contemporaneously Correlated," *Journal of the American Statistical Association*, 62, 500-509.

Searle S.R. (1971), "Topics in Variance Component Estimation," *Biometrics*, 26, 1-76.

Seely, J. (1969), "Estimation in Finite-Dimensional Vector Spaces with Application to the Mixed Linear Model," Ph.D. dissertation, Department of Statistics, Iowa State University.

Seely, J. (1970a), "Linear Spaces and Unbiased Estimation," *Annals of Mathematical Statistics*, 41, 1725-1734.

Seely, J. (1970b), "Linear Spaces and Unbiased Estimation - Application to the Mixed Linear Model," *Annals of Mathematical Statistics*, 41, 1735-1748.

Seely, J. and Soong, S. (1971), "A Note on MINQUE's and Quadratic Estimability," Corvallis, Oregon: Oregon State University.

Seely, J. and Zyskind, G. (1971), "Linear Spaces and Minimum Variance Unbiased Estimation," *Annals of Mathematical Statistics*, 42, 691-703.

Theil, H. (1961), *Economic Forecasts and Policy*, Second Edition, Amsterdam: North-Holland, 435-437.

Wansbeek, T., and Kapteyn, Arie (1989), "Estimation of the Error-Components Model with Incomplete Panels," *Journal of Econometrics,* 41, 341-361.

Wu, D. M. (1973), "Alternative Tests of Independence between Stochastic Regressors and Disturbances," *Econometrica*, 41(4), 733-750.

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.

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

The Institute is a private company devoted to the support and further development of its software and related services.