THE POWER TO KNOW.

SAS/ETS® 13.2 User's Guide

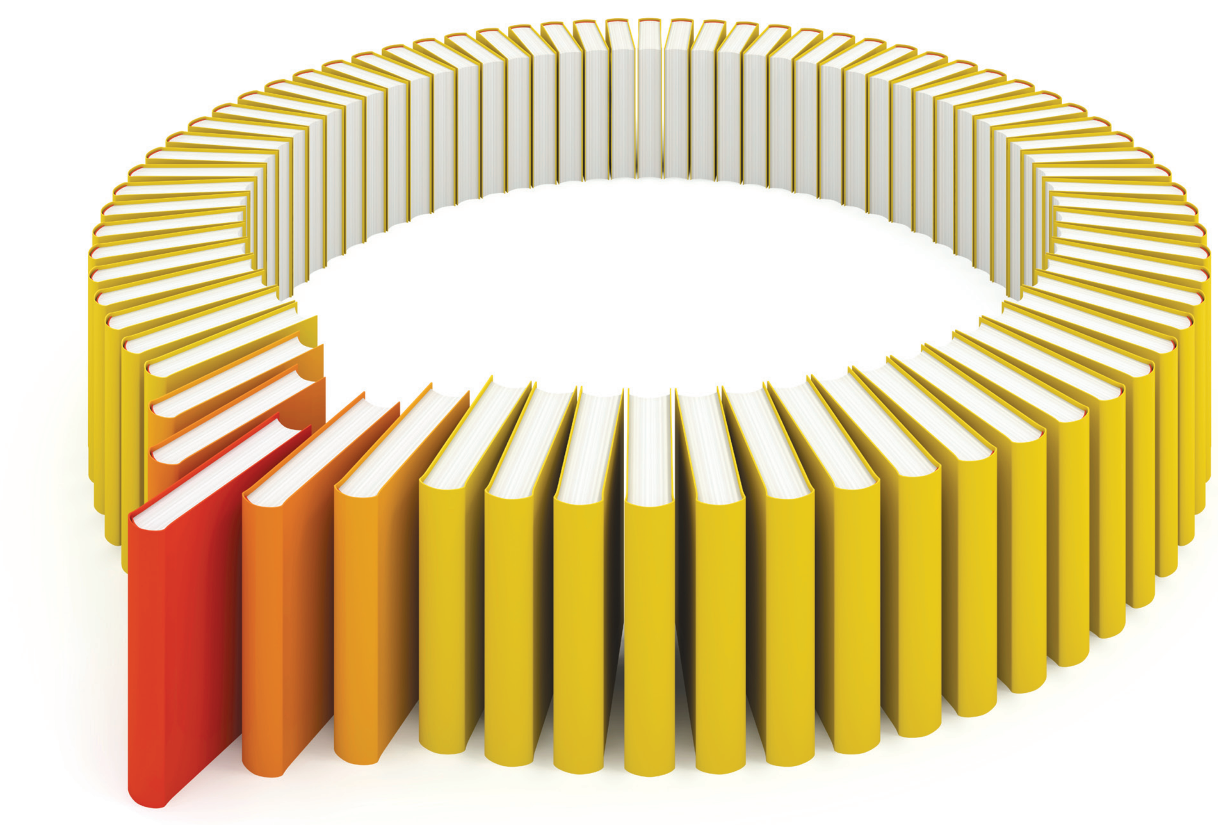
The TSCSREG Procedure

This document is an individual chapter from *SAS/ETS® 13.2 User's Guide*.

The correct bibliographic citation for the complete manual is as follows: SAS Institute Inc. 2014. *SAS/ETS® 13.2 User's Guide*. Cary, NC: SAS Institute Inc.

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

August 2014



Chapter 33
The TSCSREG Procedure

Contents

Overview: The TSCSREG Procedure

The TSCSREG (time series cross section regression) 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.

The TSCSREG procedure is very similar to the PANEL procedure; for full description, syntax details, models, and estimation methods, see Chapter 20, "The PANEL Procedure." The TSCSREG procedure is no longer being updated, and it shares the code base with the PANEL procedure.

The original TSCSREG procedure was developed by Douglas J. Drummond and A. Ronald Gallant, and contributed to the Version 5 SUGI Supplemental Library in 1979. The original code was changed substantially over the years. Additional new methods as well as other new features are currently included in the PANEL PROCEDURE. SAS Institute would like to thank Dr. Drummond and Dr. Gallant for their contribution of the original version of the TSCSREG procedure.

Getting Started: The TSCSREG Procedure

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 that contains 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 that use 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.

Specifying the Regression Model

Next, specify the linear regression model with a MODEL statement, as shown in the following statements.

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

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.

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 "Details: The TSCSREG Procedure" on page 2300 for more information about these methods and the error structures they assume.

By default, the two-way random-effects error model structure is used while Fuller-Battese and Wansbeek-Kapteyn methods are used for the estimation of variance components in balanced data and unbalanced data, respectively. Thus, the preceding example is the same as specifying the RANTWO option, as shown in the following statements:

```
proc tscsreg data=a;
   id state date;
   model y = x1 x2 / rantwo;
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 by using different options.

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-square measure, which is the most appropriate goodness-of-fit measure for models estimated by 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-square measure is equivalent to the usual R-square measure.

Estimation Techniques

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

The output from TSCSREG is identical to what one would obtain from creating dummy variables to represent the cross-sectional and time (fixed) effects. The output is presented in this manner to facilitate comparisons to the least squares dummy variables estimator (LSDV). As such, the inclusion of a intercept term implies that one dummy variable must be dropped. The actual estimation of the fixed-effects models is not LSDV. LSDV is much too cumbersome to implement. Instead, TSCSREG operates in a two step fashion. In the first step, the following occurs:

- *One-way fixed-effects model*: In the one-way fixed-effects model, the data is transformed by removing the cross-sectional means from the dependent and independent variables. The following is true:

  $$\tilde{y}_{it} = y_{it} - \bar{y}_{i\cdot}$$

  $$\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot}$$

- *Two-way fixed-effects model*: In the two-way fixed-effects model, the data is transformed by removing the cross-sectional and time means and adding back the overall means:

  $$\tilde{y}_{it} = y_{it} - \bar{y}_{i\cdot} - \bar{y}_{\cdot t} + \bar{\bar{y}}$$

  $$\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{\cdot t} + \bar{\bar{\mathbf{x}}}$$

where the symbols:

y_{it} and $\mathbf{x}_{it}$ are the dependent variable (a scalar) and the explanatory variables (a vector whose columns are the explanatory variables not including a constant), respectively

$\bar{y}_{i\cdot}$ and $\bar{\mathbf{x}}_{i\cdot}$ are cross section means

$\bar{y}_{\cdot t}$ and $\bar{\mathbf{x}}_{\cdot t}$ are time means

$\bar{\bar{y}}$ and $\bar{\bar{\mathbf{x}}}$ are the overall means

The second step consists of running OLS on the properly demeaned series, provided that the data are balanced. The unbalanced case is slightly more difficult, because the structure of the missing data must be retained. For this case, PROC TSCSREG uses a slight specialization on Wansbeek and Kapteyn.

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, σ_e^2, and 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. See to Greene (1990) for details.

```
title1;
data greene;
   input firm year output cost @@;
   df1 = firm = 1;
   df2 = firm = 2;
   df3 = firm = 3;
   df4 = firm = 4;
   df5 = firm = 5;
   d60 = year = 1960;
   d65 = year = 1965;
   d70 = year = 1970;
datalines;
    1 1955   5.36598   1.14867  1 1960   6.03787   1.45185

... more lines ...
```

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 RANTWO option specifies the two-way random-effects error model where the variance components are estimated by the Fuller-Battese method, because the data are balanced and the parameters are efficiently estimated by using the GLS method. The variance components model used by the Fuller-Battese method is

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

The following statements fit this model.

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

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

The TSCSREG procedure output is shown in Figure 33.1. A model description is printed first; it 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.

Figure 33.1 The Variance Components Estimates

The TSCSREG Procedure
Fuller and Battese Variance Components (RanTwo)

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

Variable	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-2.99992	0.6478	-4.63	0.0001
output	1	0.746596	0.0762	9.80	<.0001

Syntax: The TSCSREG Procedure

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* **;**
  TEST *equation1 < ,equation2. . . >* **;**

Functional Summary

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

Table 33.1 Functional Summary

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 two-way random-effects model	MODEL	RANTWO
specify Fuller-Battese method	MODEL	FULLER
specify PARKS	MODEL	PARKS
specify Da Silva method	MODEL	DASILVA

Description	Statement	Option
specify order of the moving-average error process for Da Silva method	MODEL	M=
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 in 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 the DATA= option, 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*
the parameter estimates. When the OUTEST= option is not specified, the OUTEST= data set is not created.

OUTCOV

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

OUTCORR

CORROUT
writes the correlation matrix of the parameter estimates to the OUTEST= data set.

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.

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 might 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. For example,

```
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 FIXONE, FIXTWO, RANONE, RANTWO, FULLER, PARKS, and DASILVA options. More than one of these options can be used, in which case the analysis is repeated for each error structure model specified.

Models can be given labels up to 32 characters in length. 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** *response = regressors / options* **;**

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 with the one-way model that corresponds to group effects only.

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 by using the Fuller-Battese method, which assumes a variance components model for the error structure.

PARKS
specifies that the model be estimated by using the Parks method, which assumes a first-order autoregressive model for the error structure.

DASILVA
specifies that the model be estimated by using the Da Silva method, which assumes a mixed variance-component moving-average model for the error structure.

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.

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

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

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 statements illustrate the use of the TEST statement:

```
proc tscsreg;
   model y = x1 x2 x3;
   test x1 = 0, x2 * .5 + 2 * x3= 0;
   test_int: test intercept=0, x3 = 0;
```

Note that a test of the following form is not permitted:

```
test_bad: test x2 / 2 + 2 * x3= 0;
```

Do not use the division sign in test/restrict statements.

Details: The TSCSREG Procedure

Models, estimators, and methods are covered in detail in Chapter 20, “The PANEL Procedure.”

ODS Table Names

PROC TSCSREG assigns a name to each table it creates. You can use these names to reference the table when you use the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in the following table.

Table 33.2 ODS Tables Produced in PROC TSCSREG

ODS Table Name	Description	Option
ODS Tables Created by the MODEL Statement		
ModelDescription	Model description	default
FitStatistics	Fit statistics	default
FixedEffectsTest	*F* test for no fixed effects	FIXONE, FIXTWO, RANONE, RANTWO
ParameterEstimates	Parameter estimates	default
CovB	Covariance of parameter estimates	COVB
CorrB	Correlations of parameter estimates	CORRB
VarianceComponents	Variance component estimates	FULLER, DASILVA, M=, RANONE, RANTWO
RandomEffectsTest	Hausman test for random effects	FULLER, DASILVA, M=, RANONE, RANTWO
AR1Estimates	First order autoregressive parameter estimates	PARKS, RHO
EstimatedPhiMatrix	Estimated phi matrix	PARKS
EstimatedAutocovariances	Estimates of autocovariances	DASILVA, M=
ODS Tables Created by the TEST Statement		
TestResults	Test results	

Examples: The TSCSREG Procedure

For examples of analysis of panel data, see Chapter 20, "The PANEL Procedure."

References

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

Subject Index

Syntax Index