SAS/ETS® 14.2 User's Guide
The PANEL Procedure

This document is an individual chapter from *SAS/ETS® 14.2 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2016. *SAS/ETS® 14.2 User's Guide*. Cary, NC: SAS Institute Inc.

SAS/ETS® 14.2 User's Guide

SAS Institute Inc., SAS Campus Drive, Cary, NC 27513-2414

November 2016

Chapter 26
The PANEL Procedure

Contents

Overview: PANEL Procedure

The PANEL procedure analyzes a class of linear econometric models that commonly arise when time series and cross-sectional data are combined. This type of pooled data on time series cross-sectional bases is often referred to as panel data. Typical examples of panel data include observations over time on households, countries, firms, trade, and so on. For example, in the case of survey data on household income, the panel is created by repeatedly surveying the same households over different time periods (years).

Regression models for panel data are characterized by an error structure that can be divided into a cross-sectional component, a time component, and an observation-level component. The panel data models can be grouped into several categories depending on the exact structure of the error term. The PANEL procedure uses the following error structures and the corresponding methods to analyze data:

- one-way and two-way models

- fixed-effects, random-effects, and hybrid models
- autoregressive models
- moving average models
- dynamic-panel models

A one-way model depends only on the cross section to which the observation belongs. A two-way model depends on both the cross section and the time period to which the observation belongs.

Apart from the possible one-way or two-way nature of the effect, the other source of disparity between the possible specifications is 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.

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 the best linear unbiased estimator. Random-effects models use a two-stage approach. In the first stage, variance components are calculated by using methods described by Fuller and Battese (1974); Wansbeek and Kapteyn (1989); Wallace and Hussain (1969); Nerlove (1971). In the second stage, variance components are used to standardize the data, and then ordinary least squares (OLS) regression is performed.

Random-effects models are more efficient than fixed-effects models, and they have the ability to estimate effects for variables that do not vary within cross sections. The cost of these added features is that random effects models carry much more stringent assumptions than their fixed-effects counterparts. The PANEL procedure supports models that blend the desirable features of both random and fixed effects. These hybrid models are those by Hausman and Taylor (1981) and Amemiya and MaCurdy (1986).

Two types of models in the PANEL procedure accommodate an autoregressive structure: the Parks method estimates a first-order autoregressive model with contemporaneous correlation, and the dynamic panel estimator estimates an autoregressive model with a lagged dependent variable as a regressor.

The Da Silva method estimates a mixed variance-component moving-average error process. The regression parameters are estimated by two-step generalized least squares (GLS).

The PANEL procedure enhances the features that were previously implemented in the TSCSREG procedure. The following list shows the most important additions:

- You can fit models for dynamic panel data by using the generalized method of moments (GMM).
- The Hausman-Taylor and Amemiya-MaCurdy estimators offer a compromise between fixed- and random-effects estimation in models where some variables are correlated with individual effects.
- The MODEL statement supports between and pooled estimation.
- The variance components for random-effects models can be calculated for both balanced and unbalanced panels by using the methods described by Fuller and Battese (1974); Wansbeek and Kapteyn (1989); Wallace and Hussain (1969); Nerlove (1971).
- The CLASS statement allows classification variables (and their interactions) directly into the analysis.
- The TEST statement performs Wald, Lagrange multiplier, and likelihood ratio tests.
- The RESTRICT statement specifies linear restrictions on the parameters.

- The FLATDATA statement processes data in compressed (wide) form.
- Several methods that produce heteroscedasticity-consistent (HCCME) and heteroscedasticity- and autocorrelation-consistent (HAC) covariance matrices are supported, because the presence of heteroscedasticity and autocorrelation can result in inefficient and biased estimates of the covariance matrix in an OLS framework.
- Tests are added for poolability, panel stationarity, the existence of cross sectional and time effects, autocorrelation, and cross-sectional dependence.
- The LAG and related statements provide functionality for creating lagged variables from within the PANEL procedure. Using these statements is preferable to using the DATA step because creating lagged variables in a panel setting can prove difficult, often requiring multiple loops and careful consideration of missing values.

  Working within the PANEL procedure makes the creation of lagged values easy. The LAG statement leaves missing values as is. Alternatively, missing values can be replaced with zeros, overall mean, time mean, or cross section mean by using the ZLAG, XLAG, SLAG, or CLAG statement, respectively.
- The OUTPUT statement enables you to output data and estimates that can be used in other analyses.
- The COMPARE statement constructs tables that enable you to easily compare parameters across multiple models and estimators.

Getting Started: PANEL Procedure

The following example uses cost function data from Greene (1990) to estimate a variance components model. The variable Production is the log of output in millions of kilowatt-hours, and Cost is the log of cost in millions of dollars. For more information, see Greene (1990).

```
data greene;
   input firm year production cost @@;
datalines;
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

... more lines ...
```

Suppose you decide to fit the following model to the data:

$$C_{it} = \text{Intercept} + \boldsymbol{\beta} P_{it} + v_i + e_t + \epsilon_{it} \quad i = 1, \ldots, N;\ t = 1, \ldots, T$$

where C_{it} and P_{it} represent the cost and production, and v_i, e_t and ϵ_{it} are the cross-sectional, time series, and error variance components.

If you assume that the time and cross-sectional effects are random, you are left with four possible estimators for the variance components. You choose Fuller-Battese.

The following statements fit this model:

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

proc panel data=greene;
      model cost = production / rantwo vcomp = fb;
      id firm year;
run;
```

The PANEL procedure output is shown in Figure 26.1. A model description is printed first, which reports the estimation method used and the number of cross sections and time periods. Fit statistics and variance components estimates are printed next. A Hausman specification test compares this model to its fixed-effects analog. Finally, the table of regression parameter estimates shows the estimates, standard errors, and *t* tests.

Figure 26.1 The Variance Components Estimates

The PANEL 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			
Coefficients	DF	m Value	Pr > m
1	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
production	1	0.746596	0.0762	9.80	<.0001

Syntax: PANEL Procedure

The following statements are used with the PANEL procedure:

PROC PANEL *options* ;
BY *variables* ;
CLASS *variables* < / *options* > ;
COMPARE < *model-list* > < / *options* > ;
FLATDATA *options* < / **OUT**=*SAS-data-set* > ;
ID *cross-section-id time-series-id* ;
INSTRUMENTS *options* ;
LAG *lag-specifications* / **OUT**=*SAS-data-set* ;
MODEL *response* = *regressors* < / *options* > ;
OUTPUT < *options* > ;
RESTRICT *equation1* < ,*equation2*. . . > ;
TEST *equation1* < ,*equation2*. . . > ;

Functional Summary

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

Table 26.1 Functional Summary

Description	Statement	Option
Data Set Options		
Includes correlations in the OUTEST= data set	PANEL	CORROUT
Includes covariances in the OUTEST= data set	PANEL	COVOUT
Specifies the input data set	PANEL	DATA=
Specifies variables to keep but not transform	FLATDATA	KEEP=
Specifies the output data set for CLASS STATEMENT	CLASS	OUT =
Specifies the output data set	FLATDATA	OUT =
Specifies the name of an output SAS data set	OUTPUT	OUT=
Writes parameter estimates to an output data set	PANEL	OUTEST=
Writes the transformed series to an output data set	PANEL	OUTTRANS=
Requests that the procedure produce graphics via the Output Delivery System	PANEL	PLOTS=
Declaring the Role of Variables		
Specifies BY-group processing	BY	
Specifies the classification variables	CLASS	
Transfers the data into uncompressed form	FLATDATA	

Table 26.1 *continued*

Description	Statement	Option
Specifies the cross section and time ID variables	ID	
Declares instrumental variables	INSTRUMENTS	
Lag Generation		
Specifies output data set for lags where missing values are replaced with the cross section mean	CLAG	OUT=
Specifies output data set for lags with missing values included	LAG	OUT=
Specifies output data set for lags where missing values are replaced with the time period mean	SLAG	OUT=
Specifies output data set for lags where missing values are replaced with overall mean	XLAG	OUT=
Specifies output data set for lags where missing values are replaced with zero	ZLAG	OUT=
Printing Control Options		
Prints correlations of the estimates	MODEL	CORRB
Prints covariances of the estimates	MODEL	COVB
Suppresses printed output	MODEL	NOPRINT
Requests that the procedure produce graphics via the Output Delivery System	MODEL	PLOTS=
Prints fixed effects	MODEL	PRINTFIXED
Performs tests of linear hypotheses	TEST	
Model Estimation Options		
Specifies the Amemiya-MaCurdy model	MODEL	AMACURDY
Requests the R_ρ statistic for serial correlation under fixed effects	MODEL	BFN
Requests the Baltagi and Li joint Lagrange multiplier (LM) test for serial correlation and random cross-sectional effects	MODEL	BL91
Requests the Baltagi and Li LM test for first-order correlation under fixed effects	MODEL	BL95
Requests the Breusch-Pagan test for one-way random effects	MODEL	BP
Requests the Breusch-Pagan test for two-way random effects	MODEL	BP2
Requests the Bera, Sosa Escudero, and Yoon modified Rao's score test	MODEL	BSY
Specifies the between-groups model	MODEL	BTWNG
Specifies the between-time-periods model	MODEL	BTWNT

Table 26.1 *continued*

Description	Statement	Option
Requests the Berenblut-Webb statistic for serial correlation under fixed effects	MODEL	BW
Requests cross-sectional dependence tests	MODEL	CDTEST
Requests the clustered HCCME estimator for the covariance matrix	MODEL	CLUSTER
Specifies the Da Silva method	MODEL	DASILVA
Requests the Durbin-Watson statistic for serial correlation under fixed effects	MODEL	DW
Specifies the one-way fixed-effects model	MODEL	FIXONE
Specifies the one-way fixed-effects model with respect to time	MODEL	FIXONETIME
Specifies the two-way fixed-effects model	MODEL	FIXTWO
Specifies the first-differenced methods for one-way models	MODEL	FDONE
Specifies the first-differenced methods for one-way models with respect to time	MODEL	FDONETIME
Specifies the first-differenced methods for two-way models	MODEL	FDTWO
Specifies the Moore-Penrose generalized inverse	MODEL	GINV=G4
Requests the Gourieroux, Holly, and Monfort test for two-way random effects	MODEL	GHM
Specifies the dynamic panel model (one-step GMM estimation)	MODEL	GMM1
Specifies the dynamic panel model (two-step GMM estimation)	MODEL	GMM2
Requests the HAC estimator for the variance-covariance matrix	MODEL	HAC=
Requests the HCCME estimator for the covariance matrix	MODEL	HCCME=
Requests the Honda test for one-way random effects	MODEL	HONDA
Requests the Honda test for two-way random effects	MODEL	HONDA2
Specifies the Hausman-Taylor model	MODEL	HTAYLOR
Specifies the dynamic panel estimator model (iterated GMM)	MODEL	ITGMM
Requests the King and Wu test for two-way random effects	MODEL	KW
Specifies the order of the moving average error process for Da Silva method	MODEL	M=
Suppresses the intercept term	MODEL	NOINT
Specifies the Parks method	MODEL	PARKS
Prints the Φ matrix for Parks method	MODEL	PHI

Table 26.1 *continued*

Description	Statement	Option
Specifies the pooled model	MODEL	POOLED
Requests poolability tests for one-way fixed effects and pooled model	MODEL	POOLTEST
Specifies the one-way random-effects model	MODEL	RANONE
Specifies the two-way random-effects model	MODEL	RANTWO
Prints autocorrelation coefficients for Parks method	MODEL	RHO
Controls the check for singularity	MODEL	SINGULAR=
Specifies the method for panel unit root/stationarity test	MODEL	UROOTTEST=
Specifies the method for the variance components estimator	MODEL	VCOMP=
Specifies linear equality restrictions on the parameters	RESTRICT	
Specifies the TEST statement	TEST	WALD, LM, LR
Requests the Wooldridge (2002) test for the presence of unobserved effects	MODEL	WOOLDRIDGE02
Comparing Models		
Create tables with side-by-side model comparisons	COMPARE	

PROC PANEL Statement

PROC PANEL *options* **;**

The following options can be specified in the PROC PANEL 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.

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" on page 1892 for details about the structure of the OUTEST= data set.

OUTTRANS=*SAS-data-set*
names an output data set to contain the transformed data. Several models supported by the PANEL procedure are estimated by first transforming the data and then applying standard regression techniques to the transformed data. This option allows you access to the transformed data. See the section "OUTTRANS= Data Set" on page 1893 for details about the structure of the OUTTRANS= data set.

OUTCOV

COVOUT

writes the standard errors and covariance matrix of the parameter estimates to the OUTEST= data set. See the section "OUTEST= Data Set" on page 1892 for details.

OUTCORR

CORROUT

writes the correlation matrix of the parameter estimates to the OUTEST= data set. See the section "OUTEST= Data Set" on page 1892 for details.

PLOTS < (*global-plot-options* **< (NCROSS=***value***) >) > < = (***specific-plot-options***) >**

selects plots to be produced via the Output Delivery System. For general information about ODS Graphics, see Chapter 21, "Statistical Graphics Using ODS" (*SAS/STAT User's Guide*). The *global-plot-options* apply to all relevant plots generated by the PANEL procedure.

Global Plot Options

The following *global-plot-options* are supported:

ONLY

suppresses the default plots. Only the plots specifically requested are produced.

UNPACKPANEL

UNPACK

displays each graph separately. By default, some graphs can appear together in a single panel.

NCROSS=*value*

specifies the number of cross sections to be combined into one time series plot.

Specific Plot Options

The following *specific-plot-options* are supported:

ACTSURFACE	produces a surface plot of actual values.
ALL	produces all appropriate plots.
FITPLOT	plots the predicted and actual values.
NONE	suppresses all plots.
PREDSURFACE	produces a surface plot of predicted values.
QQ	produces a Q-Q plot of residuals.
RESIDSTACK | RESSTACK	produces a stacked plot of residuals.
RESIDSURFACE	produces a surface plot of residual values.
RESIDUAL | RES	plots the residuals.
RESIDUALHISTOGRAM | RESIDHISTOGRAM	plots the histogram of residuals.

For more details, see the section "Creating ODS Graphics" on page 1891.

In addition, any of the following MODEL statement options can be specified in the PROC PANEL statement: CORRB, COVB, FIXONE, FIXONETIME, FIXTWO, FDONE, FDONETIME, FDTWO, BTWNG, BTWNT, POOLED, RANONE, RANTWO, HTAYLOR, AMACURDY, PARKS, DASILVA, NOINT, NOPRINT, PRINTFIXED, M=, PHI, RHO, VCOMP=, and SINGULAR=. When specified in the PROC PANEL statement, these options apply globally to every MODEL statement. See the section "MODEL Statement" on page 1806 for a complete description of each of these options.

BY Statement

BY *variables* **;**

A BY statement obtains separate analyses on observations in groups that are defined by the BY variables. When a BY statement appears, the input data set must be sorted both by the BY variables and by cross section and time period within the BY groups.

The following statements show an example:

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

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

CLASS Statement

CLASS *variables* < / **OUT=***SAS-data-set*> **;**

The CLASS statement names the classification variables to be used in the analysis. Classification variables can be either character or numeric.

The OUT=*SAS-data=set* option enables you to output the regression dummy variables used to represent the classification variables, augmented by a copy of the original data.

COMPARE Statement

COMPARE < *model-list*> < / *options*> **;**

A COMPARE statement creates tables of side-by-side comparisons of parameter estimates and other model statistics. You can fit multiple models simultaneously by specifying multiple MODEL statements, and you can specify a COMPARE statement to create tables that compare the models.

The COMPARE statement creates two tables: the first table compares model fit statistics such as R-square and mean square error; the second table compares regression coefficients, their standard errors, and (optionally) *t* tests.

By default, comparison tables are created for all fitted models, but you can use the optional *model-list* to limit the comparison to a subset of the fitted models. The *model-list* consists of a set of model labels, as specified in the MODEL statement; for more information see the section "MODEL Statement" on page 1806. If a model does not have a label, you refer to it generically as "Model *i*," where the corresponding model is the *i*th MODEL statement specified. If model labels are longer than 16 characters, then only the first 16 characters of the labels in *model-list* are used to determine a match.

You can specify one or more COMPARE statements. The following code illustrates the use of the COMPARE statement:

```
proc panel data=a;
   id csid tsid;
   mod_one: model y = x1 x2 x3    / fixone;
   model "Second Model" y = x1 x2 / fixone;
   model y = x1 x2 x3 x4          / fixone;
   compare;
   compare "Second Model" "Model 3";
run;
```

The first COMPARE statement compares all three fitted models. The second COMPARE statement compares the second and third models and uses the generic "Model 3" to identify the third model.

You can specify the following *options* in the COMPARE statement after a slash (/):

MSTAT(*mstat-list***)**
specifies a list of model fit statistics to be displayed. A set of statistics is displayed by default, but you can use this option to specify a custom set of model statistics.

mstat-list can contain one or more of the following keywords:

ALL
displays all model fit statistics. Not all statistics are appropriate for all models, and thus not always calculated. A blank cell in the table indicates that the statistic is not appropriate for that model.

DFE
displays the error degrees of freedom. This statistic is displayed by default.

F
displays the *F* statistic of the overall test for no fixed effects.

FNUMDF
displays the numerator degrees of freedom of the overall test for no fixed effects.

FDENDF
displays the denominator degrees of freedom of the overall test for no fixed effects.

M
displays the Hausman test *m* statistic.

MDF
displays the Hausman test degrees of freedom.

MSE
displays the model mean square error. This statistic is displayed by default.

NCS
displays the number of cross sections. This statistic is displayed by default.

NONE
suppresses the table of model fit statistics.

NTS
displays the maximum time-series length. This statistic is displayed by default.

PROBF
displays the significance level of the overall test for no fixed effects.

PROBM
displays the significance level of the Hausman test.

RMSE
displays the model root mean square error.

RSQUARE
displays the model R-square fit statistic. This statistic is displayed by default.

SSE
displays the model sum of squares.

VARCS
displays the variance component due to cross sections in random-effects models.

VARERR
displays the variance component due to error in random-effects models.

VARTS
displays the variance component due to time series in random-effects models.

OUTPARM=*SAS-data-set*
names an output data set to contain the data from the comparison table for parameter estimates, standard errors, and *t* tests.

OUTSTAT=*SAS-data-set*
names an output data set to contain the data from the comparison table for model fit statistics, such as R-square and mean square error.

PSTAT(*pstat-list***)**
specifies a list of parameter statistics to be displayed. By default, estimated regression coefficients and their standard errors are displayed. Use this option to specify a custom set of parameter statistics.

pstat-list can contain one or more of the following keywords:

ALL
displays all parameter statistics.

ESTIMATE
displays the estimated regression coefficient. This statistic is displayed by default.

NONE
suppresses the table of parameter statistics.

STDERR
displays the standard error. This statistic is displayed by default.

PROBT
displays the significance level of the *t* test.

T
displays the *t* statistic.

See Example 26.4 for a demonstration of the COMPARE statement.

FLATDATA Statement

FLATDATA *options* < /**OUT=***SAS-data-set* > **;**

The FLATDATA statement allows you to use PROC PANEL when you have data in flat (or wide) format, where all measurements for a given cross section are contained within one observation. See Example 26.7 for a demonstration. If you have flat data, you should issue the FLATDATA statement first in PROC PANEL, before you reference any variables you create with this statement.

The following options must be specified in the FLATDATA statement:

BASE=(*basename basename . . . basename***)**
specifies the variables that are to be transformed into a proper PROC PANEL format. All variables to be transformed must be named according to the convention: *basename*_timeperiod. You supply just the base names, and the procedure extracts the appropriate variables to transform. If some year's data are missing for a variable, then PROC PANEL detects this and fills in with missing values.

INDID=*variable*
names the variable in the input data set that uniquely identifies each individual. The INDID variable can be a character or numeric variable.

KEEP=(*variable variable . . . variable***)**
specifies the variables that are to be copied without any transformation. These variables remain constant with respect to time when the data are converted to PROC PANEL format. This is an optional item.

TSNAME=*name*
specifies a name for the generated time identifier. The name must satisfy the requirements for the name of a SAS variable. The name can be quoted, but it must not be the name of a variable in the input data set.

The following options can be specified on the FLATDATA statement after the slash (/):

OUT=*SAS-data-set*
saves the converted flat data set to a PROC PANEL formatted data set.

ID Statement

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

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 PANEL 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 PANEL 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 to sort the input data set with a BY statement with the variables listed exactly as they are listed in the ID statement, as shown in the following:

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

proc panel data=a;
   id csid tsid;
   ...
run;
```

INSTRUMENTS Statement

INSTRUMENTS *options* **;**

The INSTRUMENTS statement selects variables to be used in the moment condition equations of the dynamic panel estimator. It is also used to specify variables that are correlated with individual effects during Hausman-Taylor (HTAYLOR) or Amemiya-MaCurdy (AMACURDY) estimation.

You can specify the following *options*:

CONSTANT
includes an intercept (column of ones) as an uncorrelated exogenous instrument.

CORRELATED=(*variable variable . . . variable***)**
specifies a list of variables correlated with the unobserved individual effects. These variables are correlated with the error terms in the level equations, so they are not used in forming moment conditions from those equations.

DEPVAR<(LEVEL | DIFF | DIFFERENCE | BOTH)>
specifies instruments related to the dependent variable. With LEVEL, the lagged dependent variables are included as instruments for differenced equations. With DIFFERENCE, the differenced dependent variable is included as instruments for equations. With BOTH or nothing specified, both level and differenced dependent variables are included in the instrument matrix.

DIFFEQ=(*variable variable . . . variable***)** *or equivalently* **DIFFERENCEDEQ=(***variable . . . variable***)**
specifies a list of variables that can be used as standard instruments for the differenced equations.

EXOGENOUS=(*variable variable . . . variable***)**
specifies a list of variables that are not correlated with the disturbances given the unobserved individual effects.

LEVELEQ=(*variable variable . . . variable***)** *or equivalently* **LEVELSEQ=(***variable . . . variable***)**
specifies a list of variables that can be used as standard instruments for the level equations.

PREDETERMINED=(*variable variable . . . variable***)**
specifies a list of variables whose future realizations can be correlated with the disturbances but whose present and past realizations are not conditional on the individual effects.

For estimation with dynamic panels, a variable can be used as an instrument only if it is either exogenous or predetermined, therefore the variables listed in the CORRELATED= option must be included in either the EXOGENOUS= list or the PREDETERMINED= list. If a variable listed in the EXOGENOUS= list is not included in the CORRELATED= list, then it is considered to be uncorrelated to the error term in the level equations, which consist only of the individual effects and the disturbances. Moreover, it is uncorrelated with the error term in the differenced equations, which consist only of the disturbances. For example, in the following statements, the exogenous instruments are Z1, Z2, and X1. Because Z1 is an instrument that is correlated with the individual fixed effects, it is included in the differenced equations but not in the level equations. Because Z2 is not correlated with either the individual effects or the disturbances, it is included in both the level equations and the differenced equations.

```
proc panel data=a;
   instruments exogenous = (Z1 Z2 X1) correlated = (Z1) constant depvar;
   model Y = X1 X2 X3 / gmm1;
run;
```

For a detailed discussion of the model set up and the use of the INSTRUMENTS statement for dynamic panels, see "Dynamic Panel Estimators" on page 1852.

For Hausman-Taylor or Amemiya-MaCurdy estimation, you specify which variables are correlated with the individual effects by using the CORRELATED= option. All other options are ignored. For these estimators, the specified variables are not instruments—they are merely designated as correlated. The instruments are determined by the method; for more information, see the section "Hausman-Taylor Estimation" on page 1846.

```
proc panel data=a;
   instruments correlated = (X2 Z2);
   model Y = X1 X2 Z1 Z2 / htaylor;
run;
```

An INSTRUMENT statement is required for each MODEL statement. For example, if there are two models to be estimated by using GMM1 within one PANEL procedure, then there should be two INSTRUMENT statements, as follows:

```
proc panel data=test;
  id cs ts;
  instruments depvar predetermined = (x1 x2)
                     exogenous     = (x3 x4 x5)
                     correlated    = (x3 x4 x5);
  model y = y_1 x1 x2 / gmm1;
  instruments        predetermined = (x2 x4)
                     exogenous     = (x3 x5)
                     correlated    = (x3 x4);
  model y = y_1 x2 / gmm1;
run;
```

LAG, CLAG, SLAG, XLAG, and ZLAG Statements

LAG $var_1(lag_1\ lag_2\ \ldots lag_T)\ \ldots var_N(lag_1\ lag_2\ \ldots lag_T)$ / **OUT**= *SAS-data-set* **;**

Generally, creating lags of variables in a panel setting is a tedious process requiring many DATA step statements. The PANEL procedure enables you to generate lags of any series without stepping through individual time series. The LAG statement is a data set generation tool. You can specify more than one LAG statement. Analyzing the generated lagged data requires a subsequent call to PROC PANEL.

The OUT= option is required. The output data set includes all variables in the input set, plus the generated lags, named using the convention *varname_lag*. The LAG statement tends to generate many missing values in the data. This can be problematic because the number of usable observations diminishes with the lag length. Therefore, PROC PANEL offers the following alternatives to the LAG statement. The following statements can be used in place of LAG with otherwise identical syntax:

CLAG replaces missing values with the cross section mean for that variable.

SLAG replaces missing values with the time mean for that variable.

XLAG replaces missing values with the overall mean for that variable.

ZLAG replaces missing values with 0 for that variable.

For all of the above, missing values are replaced only if they are in the generated (lagged) series. Missing variables in the original variables remain unaltered.

Assume that data set A has been sorted by cross section and by time period within cross section and that the variables are Y, X1, X2, and X3. The following PROC PANEL statements generate a series with lags 1 and 3 of the X1 variable; lags 3, 6, and 9 of the X2 variable; and lag 2 of the X3 variable:

```
proc panel data=A;
   id i t;
   lag X1(1 3) X2(3 6 9) X3(2) / out=A_lag;
run;
```

If you want zeroing instead of missing values, then use ZLAG in place of LAG.

```
proc panel data=A;
   id i t;
   zlag X1(1 3) X2(3 6 9) X3(2) / out=A_zlag;
run;
```

Similarly, you can specify XLAG to replace with overall means, SLAG to replace with time means, and CLAG to replace with cross section means.

MODEL Statement

MODEL < *"string"*> *response* = *regressors* < / *options*> **;**

The MODEL statement specifies the regression model, the error structure that is assumed for the regression residuals, and the estimation technique to be used. The response variable (*response*) on the left side of the equal sign is regressed on the independent variables (*regressors*), which are listed after the equal sign. You can specify any number of MODEL statements. For each MODEL statement, you can specify only one *response*.

You can label models. Model labels are used in the printed output to identify the results for different models. If you do not specify a label, the model is referred to by numerical order wherever necessary. You can label the models in two ways:

First, you can prefix the MODEL statement by a label followed by a colon. For example:

label: **MODEL** . . . ;

Second, you can add a quoted string after the MODEL keyword. For example:

MODEL "*label*" . . . ;

Quoted-string labels are preferable because they allow spaces and special characters and because these labels are case-sensitive. If you specify both types of label, PROC PANEL uses the quoted string.

The MODEL statement supports a multitude of options, some more specific than others. Table 26.2 summarizes the *options* available in the MODEL statement. These are subsequently discussed in detail in the order in which they are presented in the table.

Table 26.2 Summary of MODEL Statement Options

Option	Description
Estimation Technique Options	
AMACURDY	Fits a one-way model by using the Amemiya-MaCurdy estimator
BTWNG	Fits the between-groups model

Table 26.2 *continued*

Option	Description
BTWNT	Fits the between-time-periods model
DASILVA	Fits a moving average model by using the Da Silva method
FDONE	Fits a one-way model by using first differences
FDONETIME	Fits a one-way model for time effects by using first differences
FDTWO	Fits a two-way model by using first differences
FIXONE	Fits a one-way fixed-effects model
FIXONETIME	Fits a one-way fixed-effects model for time effects
FIXTWO	Fits a two-way fixed effects model
GMM1	Fits a dynamic-panel model by using the one-step generalized method of moments (GMM)
GMM2	Fits a dynamic-panel model by using two-step GMM
HTAYLOR	Fits a one-way model by using the Hausman-Taylor estimator
ITGMM	Fits a dynamic-panel model by using iterated GMM
PARKS	Fits an autoregressive model by using the Parks method
POOLED	Fits the pooled regression model
RANONE	Fits a one-way random-effects model
RANTWO	Fits a two-way random-effects model
Estimation Control Options	
M=	Specifies the moving average order
NOESTIM	Limits estimation to only transforming the data
NOINT	Suppresses the intercept
SINGULAR=	Specifies a matrix inverse singularity criterion
VCOMP=	Specifies the type of variance component estimation for random-effects estimation
Dynamic Panel Estimation Control Options	
ARTEST=	Specifies the maximum order of the auto regression (AR) test
ATOL=	Specifies the convergence criterion of iterated GMM, with respect to the weighting matrix
BANDOPT=	Specifies which neighboring observations to use as instruments, whether TRAILING, CENTERED, or LEADING
BIASCORRECTED	Requests bias-corrected variances for two-step GMM
BTOL=	Specifies the convergence criterion of iterated GMM, with respect to the parameter matrix
GINV=	Specifies the type of generalized matrix inverse
MAXBAND=	Specifies the moment condition bandwidth

Table 26.2 *continued*

Option	Description
MAXITER=	Specifies the maximum iterations for iterative GMM
NODIFFS	Estimates without moment conditions from the difference equations
NOLEVELS	Estimates without moment conditions from the level equations
ROBUST	Specifies the robust covariance matrix
TIME	Includes time dummy variables in the model
Alternative Variances Options	
CLUSTER	Corrects covariance for intracluster correlation
HAC(*options*)	Specifies a heteroscedasticity- and autocorrelation-consistent (HAC) covariance
HCCME=	Specifies a heteroscedasticity-corrected covariance matrix estimator (HCCME)
NEWEYWEST(*options*)	Specifies the Newey-West covariance, a special case of the HAC covariance
Unit Root Test Options	
UROOTTEST(*test-options*)	Requests one or more panel data unit root and stationarity tests; specify *test-options* ALL through lLC within this option
STATIONARITY(*test-options*)	Synonym for UROOTTEST
ALL	Requests that all unit root tests be performed
BREITUNG(*options*)	Specifies Breitung's tests that are robust to cross-sectional dependence
COMBINATION(*options*)	Specifies one or more unit root tests that combine over all cross sections
FISHER(*options*)	Synonym for COMBINATION
HADRI(*options*)	Specifies Hadri's (2000) stationarity test
HT	Specifies the Harris and Tzavalis (1999) panel unit root test
IPS(*options*)	Specifies the Im, Pesaran, and Shin (2003) panel unit root test
LLC(*options*)	Specifies the Levin, Lin, and Chu (2002) panel unit root test
Model Specification Test Options	
BFN	Requests the R_ρ statistic for serial correlation under fixed effects
BL91	Requests the Baltagi and Li (1991) Lagrange multiplier (LM) test for serial correlation and random effects
BL95	Requests the Baltagi and Li (1995) LM test for first-order correlation under fixed effects
BP	Requests the Breusch-Pagan one-way test for random effects

Table 26.2 *continued*

Option	Description
BP2	Requests the Breusch-Pagan two-way test for random effects
BSY	Requests the Bera, Sosa Escudero, and Yoon modified Rao's score test
BW	Requests the Berenblut-Webb statistic for serial correlation under fixed effects
CDTEST(*options*)	Requests a battery of cross-sectional dependence tests.
DW	Requests the Durbin-Watson statistic for serial correlation under fixed effects
GHM	Requests the Gourieroux, Holly, and Monfort test for two-way random effects
HONDA	Requests the Honda one-way test for random effects
HONDA2	Requests the Honda two-way test for random effects
KW	Requests the King and Wu two-way test for random effects
POOLTEST	Requests poolability tests for one-way fixed effects and pooled models
WOOLDRIDGE02	Requests the Wooldridge (2002) test for unobserved effects
Printed Output Options	
CORR	Prints the parameter correlation matrix
CORRB	Synonym for CORR
COVB	Prints the parameter covariance matrix
ITPRINT	Prints the iteration history
NOPRINT	Suppress normally printed output
PHI	Prints the Φ covariance matrix for the Parks method
PRINTFIXED	Estimates and prints the fixed effects
RHO	Prints the autocorrelation coefficients for the Parks method
VAR	Synonym for COVB

You can specify the following *options* in the MODEL statement after a slash (/).

Estimation Technique Options

These options specify the assumed error structure and estimation method. You can specify more than one option, in which case the analysis is repeated for each. The default is RANTWO (two-way random effects).

All estimation methods are detailed in the section "Details: PANEL Procedure" and its subsections.

AMACURDY
requests Amemiya-MaCurdy estimation for a model that has correlated individual (cross-sectional) effects. This option requires that you specify the CORRELATED= option in the INSTRUMENTS statement.

BTWNG
estimates a between-groups model.

BTWNT
estimates a between-time-periods model.

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

FDONE
estimates a one-way model by using first-differenced methods.

FDONETIME
estimates a one-way model that corresponds to time effects by using first-differenced methods.

FDTWO
estimates a two-way model by using first-differenced methods.

FIXONE
estimates a one-way fixed-effects model that corresponds to cross-sectional effects only.

FIXONETIME
estimates a one-way fixed-effects model that corresponds to time effects only.

FIXTWO
estimates a two-way fixed-effects model.

GMM1
estimates the model in a single step by using the dynamic panel estimator method, which allows for autoregressive processes. This option requires you to specify the INSTRUMENTS statement.

GMM2
estimates the model in two steps by using the dynamic panel estimator method. The first step forms an estimator for the weighting matrix that is used in the second step. This option requires you to specify the INSTRUMENTS statement.

HTAYLOR
requests Hausman-Taylor estimation for a model that has correlated individual (cross-sectional) effects. This option requires you to specify the CORRELATED= option in the INSTRUMENTS statement.

ITGMM
estimates the model by using the dynamic panel estimator method, but requests that PROC PANEL keep updating the weighting matrix until either the parameter vector converges or the weighting matrix converges. This option requires you to specify the INSTRUMENTS statement.

PARKS
estimates the model by using the Parks method, which assumes a first-order autoregressive model for the error structure.

POOLED
estimates a pooled (OLS) model.

RANONE
estimates a one-way random-effects model.

RANTWO
estimates a two-way random-effects model.

Estimation Control Options

These options define parameters that control the estimation and can be specific to the chosen technique (for example, how to estimate variance components in a random-effects model).

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

NOESTIM
limits the estimation of a FIXONE, FIXONETIME, FDONE, FDONETIME, or RANONE model to the generation of the transformed series. This option is intended for use with an OUTTRANS= data set.

NOINT
suppresses the intercept parameter from the model.

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

VCOMP=FB | NL | WH | WK
specifies the type of variance component estimate to use. You can specify the following values:

FB	uses the Fuller and Battese method.
NL	uses the Nerlove method.
WH	uses the Wallace and Hussain method.
WK	uses the Wansbeek and Kapteyn method.

By default, VCOMP=FB for balanced data and VCOMP=WK for unbalanced data. For more information, see the sections "One-Way Random-Effects Model" on page 1838 and "Two-Way Random-Effects Model" on page 1841.

Dynamic Panel Estimation Control Options

These control options are specific to dynamic panels, where the estimation technique is specified as GMM1, GMM2, or ITGMM. For more information, see the section "Dynamic Panel Estimators" on page 1852.

ARTEST=*integer*
specifies the maximum order of the test for the presence of auto regression (AR) effects in the residual in the dynamic panel model. The value of *integer* must be between 1 and the $T-3$, inclusive, where T is the number of time periods.

ATOL=*number*
specifies the convergence criterion for the iterated generalized method of moments (GMM) when convergence of the method is determined by convergence in the weighting matrix. The convergence criterion (*number*) must be positive. If you do not specify this option, then the BTOL= option (or its default) is used.

BANDOPT=CENTERED | LEADING | TRAILING
specifies which observations are included in the instrument list when the MAXBAND= option is specified. You can specify the following values:

CENTERED uses both leading and trailing observations.

LEADING uses only leading observations.

TRAILING uses only trailing observations.

This option should be used only for exogenous instruments. By default, BANDOPT=TRAILING.

BIASCORRECTED
requests that the bias-corrected covariance matrix of the two-step dynamic panel estimator be computed. When you specify this option, the ROBUST option is disabled for the two-step GMM estimator.

BTOL=*number*
specifies the convergence criterion for iterated GMM when convergence of the method is determined by convergence in the parameter matrix. The convergence criterion (*number*) must be positive. The default is BTOL=1E–8.

GINV= G2 | G4
specifies what type of generalized inverse to use. You can specify the following values:

G2 uses the G2 generalized inverse.

G4 uses the G4 generalized inverse.

The G4 inverse is generally more stable, but numerically intensive. By default, GINV=G2.

MAXBAND=*integer*
specifies the maximum number of time periods (per instrumental variable) that are allowed into the moment condition. The acceptable range for *integer* is 1 to $T-1$, where T is the number of time periods. If BANDOPT=LEADING or CENTERED, then the default value of MAXBAND is 2. If BANDOPT=TRAILING, then the default value of MAXBAND is 1. If no BANDOPT option is specified (such as when no exogenous instruments are used), then the default value of MAXBAND is 1.

MAXITER=*integer*
specifies the maximum number of iterations for the ITGMM option. By default, MAXITER=200.

NODIFFS
estimates the dynamic panel model without moment conditions from the difference equations.

NOLEVELS
estimates the dynamic panel model without moment conditions from the level equations.

ROBUST
uses the robust weighting matrix in the calculation of the covariance matrix of the single-step, two-step, and iterated GMM dynamic panel estimators.

TIME
estimates the model by using the dynamic panel estimator method but includes time dummy variables to model any time effects in the data.

Alternative Variances Options

These options specify variance estimation other than conventional model-based variance estimation. They include the robust, cluster robust, HAC, HCCME, and Newey-West techniques.

CLUSTER
specifies the cluster correction for the covariance matrix. You can specify this option when you specify HCCME=0, 1, 2, or 3.

HAC <(*options***) >**
specifies the heteroscedasticity- and autocorrelation-consistent (HAC) covariance matrix estimator. This option is not available for between models and cannot be combined with the HCCME option.

For more information, see the section "Heteroscedasticity- and Autocorrelation-Consistent Covariance Matrices" on page 1867.

You can specify the following *options* within parentheses and separated by spaces:

BANDWIDTH=*number* | *method*
specifies the fixed bandwidth value or bandwidth selection method to be used in the kernel function. You can specify either a fixed value (*number*) or one of the *methods* shown after *number*.

number
specifies a fixed value of the bandwidth parameter.

ANDREWS91 | ANDREWS
specifies the Andrews (1991) bandwidth selection method.

NEWEYWEST94<(C=*number***)>**
NW94 <(C=*number***)>**
specifies the Newey and West (1994) bandwidth selection method. You can also specify C=*number* for the calculation of lag selection parameter; the default is C=12.

SAMPLESIZE<(*options*)>

SS<(*options*)>

calculates the bandwidth according to the following equation based on the sample size

$$b = \gamma T^r + c$$

where b is the bandwidth parameter; T is the sample size; and γ, r, and c are values specified by the following *options* within parentheses and separated by commas.

GAMMA=*number*

specifies the coefficient γ in the equation. By default, GAMMA=0.75.

RATE=*number*

specifies the growth rate r in the equation. By default, RATE=0.3333.

CONSTANT=*number*

specifies the constant c in the equation. By default, CONSTANT=0.5.

INT

specifies that the bandwidth parameter must be integer; that is, $b = \lfloor \gamma T^r + c \rfloor$, where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x.

By default, BANDWIDTH=ANDREWS91.

KERNEL=BARTLETT | PARZEN | QS | TH | TRUNCATED

specifies the type of kernel function. You can specify the following values:

BARTLETT	specifies the Bartlett kernel function.
PARZEN	specifies the Parzen kernel function.
QS	specifies the quadratic spectral kernel function.
TH	specifies the Tukey-Hanning kernel function.
TRUNCATED	specifies the truncated kernel function.

By default, KERNEL=TRUNCATED.

KERNELLB=*number*

specifies the lower bound of the kernel weight value. Any kernel weight less than *number* is regarded as 0, which accelerates the calculation in large samples, especially for the quadratic spectral kernel function. By default, KERNELLB=0.

PREWHITENING

requires prewhitening in the covariance calculation.

ADJUSTDF

requires adjustment of degrees of freedom in the covariance calculation.

HCCME= NO | *number*

specifies the type of HCCME covariance matrix. You can specify one of the following:

NO does not correct the covariance matrix.

number specifies the type of covariance adjustment. The value of *number* can be any integer from 0 to 4, inclusive.

For more information, see the section "Heteroscedasticity-Corrected Covariance Matrices" on page 1864. By default, HCCME=NO.

NEWEYWEST<(*options***)>**

specifies the well-known Newey-West estimator, a special HAC estimator that uses (1) the Bartlett kernel, (2) a bandwidth that is determined by the equation based on the sample size, $b = \lfloor \gamma T^r + c \rfloor$, and (3) no adjustment for degrees of freedom and no prewhitening. By default, the bandwidth parameter for Newey-West estimator is $\lfloor 0.75T^{0.3333} + 0.5 \rfloor$, as shown in equation (15.17) in Stock and Watson (2002). You can specify the following *options* in parentheses and separated by commas:

GAMMA=*number*

specifies the coefficient γ in the equation. By default, GAMMA=0.75.

RATE=*number*

specifies the growth rate r in the equation. By default, RATE=0.3333.

CONSTANT=*number*

specifies the constant c in the equation. By default, CONSTANT=0.5.

To specify a Newey-West bandwidth directly (and not as a function of time-series length), set GAMMA=0 and CONSTANT=*b*, where *b* is the bandwidth you want. For example, the two variance specifications in the following statements are equivalent:

```
proc panel data=A;
   id i t;
   model y = x1 x2 x3 / ranone hac(kernel = bartlett bandwidth = 3);
   model y = x1 x2 x3 / ranone neweywest(gamma = 0, constant = 3);
run;
```

Unit Root Test Options

These options request unit root tests on the dependent variable. You begin with the UROOTTEST (or its synonym STATIONARITY) option and specify everything else within parentheses after the UROOTTEST (or SINGULARITY) keyword. The BREITUNG, COMBINATION, HADRI, HT, IPS, and LLC tests are available, and you can request them all by specifying the ALL option.

UROOTTEST(*test1***<(***test-options***),** *test2***<(***test-options***)>... >** **<** *options* **>)**

STATIONARITY(*test1***<(***test-options***),** *test2***<(***test-options***)>... >** **<** *options* **>)**

specifies tests of stationarity or unit root for panel data, and specifies options for each test. These tests apply only to the dependent variable. Six tests are available: BREITUNG, COMBINATION (or FISHER), HADRI, HT, IPS, and LLC. You can specify one or more of these tests, separated by commas. You can also request all tests by specifying UROOTEST(ALL) or STATIONARITY(ALL). If you specify one or more *test-options* (separated by spaces) inside the parentheses after a particular test, they apply only to that test. If you specify one or more *options* separated by spaces after you specify the tests, they apply to all the tests. If you specify both *test-options* and *options*, the *test-options* override the *options*.

You can specify the following *tests* and *test-options*:

BREITUNG< (*test-options***) >**
performs Breitung's unbiased test, t test, and generalized least squares (GLS) t test that are robust to cross-sectional dependence. The tests are described in Breitung and Meyer (1994); Breitung (2000); Breitung and Das (2005). You can specify one or more of the following *test-options* within parentheses and separated by spaces:

DETAIL
requests that intermediate results (lag order) be printed.

LAG=*type* | *value*
specifies the method to choose the lag order for the augmented Dickey-Fuller (ADF) regressions. You can specify a *value* or one of the *types* that are shown after *value*.

value
specifies the lag order. If the lag order is too big to run linear regression (*value* > $T - k$, where T is the number of time periods and k is the number of parameters), then the lag order is set to $\left\lfloor 12(T/100)^{1/4} \right\rfloor$ or $T - k - 1$, whichever is smaller.

GS
selects the order of lags by Hall's (1994) sequential testing method, from the most general model (maximum lags) to lower order of lag terms.

SG
selects the order of lags by Hall's (1994) sequential testing method, from no lag term to maximum allowed lags.

AIC
selects the order of lags by Akaike's information criterion (AIC).

SBC
SIC
SBIC
selects the order of lags by the Bayesian information criterion (Schwarz criterion).

HQIC
selects the order of lags by the Hannan-Quinn information criterion.

MAIC
selects the order of lags by the modified AIC that is proposed by Ng and Perron (2001).

By default, LAG=MAIC.

MAXLAG=*value*
specifies the maximum lag order that the model allows. The default value is $\left\lfloor 12(T/100)^{1/4} \right\rfloor$. If *value* is larger than 0 and larger than $T - k$, then the maximum lag order is set to be the default value of $\left\lfloor 12(T/100)^{1/4} \right\rfloor$ or $T - k - 1$, whichever is smaller. This option is ignored if you specify LAG=*value*.

COMBINATION < (*test-options*) >

FISHER < (*test-options*) >

specifies combination tests that are proposed by Choi (2001); Maddala and Wu (1999). Fisher's test, as proposed by Maddala and Wu (1999), is a special case of combination tests. You can specify one or more of the following *test-options* within parentheses and separated by spaces:

TEST=ADF | PP

selects the time series unit root test for combination tests (Fisher's test). You can specify the following values:

ADF specifies the augmented Dickey-Fuller (ADF) test. The BANDWIDTH and KERNEL options are ignored because they do not pertain to ADF tests.

PP specifies the Phillips and Perron (1988) unit root test. The LAG and MAXLAG options are ignored because they do not pertain to PP tests.

By default, TEST=PP.

KERNEL=BARTLETT | PARZEN | QS | TH | TRUNCATED

specifies the type of kernel function. You can specify the following values:

BARTLETT specifies the Bartlett kernel function.

PARZEN specifies the Parzen kernel function.

QS specifies the quadratic spectral kernel function.

TH specifies the Tukey-Hanning kernel function.

TRUNCATED specifies the truncated kernel function.

By default, KERNEL=QS.

BANDWIDTH=ANDREWS | *number*

specifies the bandwidth for the kernel. You can specify one of the following:

ANDREWS selects the bandwidth by the Andrews method.

number sets the bandwidth to *number*, which must be nonnegative.

By default, BANDWIDTH=ANDREWS.

DETAIL

requests that intermediate results (lag order and long-run variance for each cross section) be printed.

LAG=*type* | *value*

specifies the method to choose the lag order for the augmented Dickey-Fuller (ADF) regressions. You can specify a *value* or one of the *types* that are shown after *value*.

value
specifies the lag order. If the lag order is too big to run linear regression (*value* > $T - k$, where T is the number of time periods and k is the number of parameters), then the lag order is set to $\left\lfloor 12(T/100)^{1/4} \right\rfloor$ or $T - k - 1$, whichever is smaller.

GS
selects the order of lags by Hall's (1994) sequential testing method, from the most general model (maximum lags) to lower order of lag terms.

SG
selects the order of lags by Hall's (1994) sequential testing method, from no lag term to maximum allowed lags.

AIC
selects the order of lags by Akaike's information criterion (AIC).

SBC
SIC
SBIC
selects the order of lags by the Bayesian information criterion (Schwarz criterion).

HQIC
selects the order of lags by the Hannan-Quinn information criterion.

MAIC
selects the order of lags by the modified AIC that is proposed by Ng and Perron (2001).

By default, LAG=MAIC.

MAXLAG=*value*
specifies the maximum lag order that the model allows. The default value is $\left\lfloor 12(T/100)^{1/4} \right\rfloor$. If *value* is larger than 0 and larger than $T - k$, then the maximum lag order is set to be the default value of $\left\lfloor 12(T/100)^{1/4} \right\rfloor$ or $T - k - 1$, whichever is smaller. This option is ignored if you specify LAG=*value*.

HADRI < (*test-options***) >**
specifies Hadri's (2000) panel stationarity test. You can specify the following *test-options*:

DETAIL
requests that intermediate results (lag order and long-run variance for each cross section) be printed.

KERNEL=BARTLETT | PARZEN | QS | TH | TRUNCATED
specifies the type of kernel function. You can specify the following values:

BARTLETT	specifies the Bartlett kernel function.
PARZEN	specifies the Parzen kernel function.
QS	specifies the quadratic spectral kernel function.
TH	specifies the Tukey-Hanning kernel function.

TRUNCATED specifies the truncated kernel function.

By default, KERNEL=QS.

BANDWIDTH=ANDREWS | *number*
specifies the bandwidth for the kernel. You can specify one of the following:

ANDREWS selects the bandwidth by the Andrews method.

number sets the bandwidth to *number*, which must be nonnegative.

By default, BANDWIDTH=ANDREWS.

HT
specifies the Harris and Tzavalis (1999) panel unit root test. No options are available for this test.

IPS < (*test-options***) >**
specifies the Im, Pesaran, and Shin (2003) panel unit root test. You can specify one or more of the following *test-options* within parentheses and separated by spaces:

DETAIL
requests that intermediate results (lag order) be printed.

LAG=*type* | *value*
specifies the method to choose the lag order for the augmented Dickey-Fuller (ADF) regressions. You can specify a *value* or one of the *types* that are shown after *value*.

value
specifies the lag order. If the lag order is too big to run linear regression (*value* > $T - k$, where T is the number of time periods and k is the number of parameters), then the lag order is set to $\left\lfloor 12(T/100)^{1/4} \right\rfloor$ or $T - k - 1$, whichever is smaller.

GS
selects the order of lags by Hall's (1994) sequential testing method, from the most general model (maximum lags) to lower order of lag terms.

SG
selects the order of lags by Hall's (1994) sequential testing method, from no lag term to maximum allowed lags.

AIC
selects the order of lags by Akaike's information criterion (AIC).

SBC

SIC

SBIC
selects the order of lags by the Bayesian information criterion (Schwarz criterion).

HQIC
selects the order of lags by the Hannan-Quinn information criterion.

MAIC
selects the order of lags by the modified AIC that is proposed by Ng and Perron (2001).

By default, LAG=MAIC.

MAXLAG=*value*
specifies the maximum lag order that the model allows. The default value is $\left\lfloor 12(T/100)^{1/4} \right\rfloor$. If *value* is larger than 0 and larger than $T-k$, then the maximum lag order is set to be the default value of $\left\lfloor 12(T/100)^{1/4} \right\rfloor$ or $T-k-1$, whichever is smaller. This option is ignored if you specify LAG=*value*.

LLC < (*test-options***) >**
specifies the Levin, Lin, and Chu (2002) panel unit root test. You can specify one or more of the following *test-options* within parentheses and separated by spaces:

DETAIL
requests that intermediate results (lag order and long-run variance for each cross section) be printed.

KERNEL=BARTLETT | PARZEN | QS | TH | TRUNCATED
specifies the type of kernel function. You can specify the following values:

BARTLETT	specifies the Bartlett kernel function.
PARZEN	specifies the Parzen kernel function.
QS	specifies the quadratic spectral kernel function.
TH	specifies the Tukey-Hanning kernel function.
TRUNCATED	specifies the truncated kernel function.

By default, KERNEL=QS.

BANDWIDTH=ANDREWS | *number*
specifies the bandwidth for the kernel. You can specify one of the following:

ANDREWS	selects the bandwidth by the Andrews method.
number	sets the bandwidth to *number*, which must be nonnegative. By default, BANDWIDTH=ANDREWS.

LAG=*type* **|** *value*
specifies the method to choose the lag order for the augmented Dickey-Fuller (ADF) regressions. You can specify a *value* or one of the *types* that are shown after *value*.

value
specifies the lag order. If the lag order is too big to run linear regression (*value* > $T-k$, where T is the number of time periods and k is the number of parameters), then the lag order is set to $\left\lfloor 12(T/100)^{1/4} \right\rfloor$ or $T-k-1$, whichever is smaller.

GS
selects the order of lags by Hall's (1994) sequential testing method, from the most general model (maximum lags) to lower order of lag terms.

SG
selects the order of lags by Hall's (1994) sequential testing method, from no lag term to maximum allowed lags.

AIC
selects the order of lags by Akaike's information criterion (AIC).

SBC
SIC
SBIC
selects the order of lags by the Bayesian information criterion (Schwarz criterion).

HQIC
selects the order of lags by the Hannan-Quinn information criterion.

MAIC
selects the order of lags by the modified AIC that is proposed by Ng and Perron (2001).

By default, LAG=MAIC.

MAXLAG=*value*
specifies the maximum lag order that the model allows. The default value is $\left\lfloor 12(T/100)^{1/4} \right\rfloor$. If *value* is larger than 0 and larger than $T-k$, then the maximum lag order is set to be the default value of $\left\lfloor 12(T/100)^{1/4} \right\rfloor$ or $T-k-1$, whichever is smaller. This option is ignored if you specify LAG=*value*.

Consider the following example, which requests two tests (LLC and BREITUNG) on the dependent variable:

```
proc panel data=A;
   id i t;
   model y = x1 x2 x3 / unitroot(llc(kernel = parzen lag = aic),
                                 breitung(lag = gs)
                                 maxlag = 2
                                 kernel = bartlett);
run;
```

For the LLC test, the lag order is selected by AIC with maximum lag order 2 and the kernel is specified as Parzen (overriding Bartlett). For the BREITUNG test, the lag order is GS with a maximum lag order 2. The KERNEL option is ignored by BREITUNG because it is not relevant to that test.

Model Specification Test Options

These options request model specification tests, such as a test for poolability in one-way models. These tests depend on the model specifications of dependent and independent variables, but not on the estimation technique that is used to fit the model. For example, a one-way test for random effects does not require you to fit a random effects model, or even a one-way model for that matter. The model fits that are required for the selected tests are performed internally.

BFN (Experimental)
requests the R_ρ statistic for serial correlation under cross-sectional fixed effects.

BL91
requests the Baltagi and Li (1991) joint Lagrange multiplier (LM) test for serial correlation and random cross-sectional effects.

BL95
requests the Baltagi and Li (1995) LM test for first-order correlation under fixed effects.

BP
requests the Breusch-Pagan one-way test for random effects.

BP2
requests the Breusch-Pagan two-way test for random effects.

BSY
requests the Bera, Sosa Escudero, and Yoon modified Rao's score test for random cross-sectional effects or serial correlation or both.

BW (Experimental)
requests the Berenblut-Webb statistic for serial correlation under cross-sectional fixed effects.

CDTEST < (P=*value*) >
requests cross-sectional dependence tests. These include the Breusch and Pagan (1980) LM test, the scaled version of the Breusch and Pagan (1980) test, and the Pesaran (2004) CD test. When you specify P=*value*, the CD test for local cross-sectional dependence is performed with the order *value*, where *value* is an integer greater than 0.

DW (Experimental)
requests the Durbin-Watson statistic for serial correlation under cross-sectional fixed effects.

GHM (Experimental)
requests the Gourieroux, Holly, and Monfort two-way test for random effects.

HONDA
requests the Honda one-way test for random effects.

HONDA2
requests the Honda two-way test for random effects.

KW
requests the King and Wu two-way test for random effects.

POOLTEST
requests poolability tests for one-way fixed effects and pooled models.

WOOLDRIDGE02
requests the Wooldridge (2002) test for the presence of unobserved effects.

Printed Output Options

These options alter how results are presented.

CORRB

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

COVB

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

ITPRINT
prints out the iteration history of the parameter and transformed sum of squared errors.

NOPRINT
suppresses the normal printed output.

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 specified. For more information, see the section "Parks Method (Autoregressive Model)" on page 1848.

PRINTFIXED
estimates and prints the fixed effects in models where they would normally be absorbed within the estimation.

RHO
prints the estimated autocorrelation coefficients for the Parks method.

OUTPUT Statement

OUTPUT < *options* > **;**

The OUTPUT statement creates an output SAS data set as specified by the following options:

OUT=*SAS-data-set*
names the output SAS data set to contain the predicted and transformed values. If the OUT= option is not specified, the new data set is named according to the DATA*n* convention.

PREDICTED=*name*

P=*name*
writes the predicted values to the output data set.

RESIDUAL=*name*

R=*name*
writes the residuals from the predicted values based on both the structural and time series parts of the model to the output data set.

RESTRICT Statement

RESTRICT < *"string"*> *equation* < *,equation2. . .* > **;**

The RESTRICT statement specifies linear equality restrictions on the parameters in the previous model statement. There can be as many unique restrictions as the number of parameters in the preceding model statement. Multiple RESTRICT statements are understood as joint restrictions on a model's parameters. Restrictions on the intercept are obtained by the use of the keyword INTERCEPT. RESTRICT statements before the first MODEL statement are automatically associated with the first MODEL statement, in addition to any RESTRICT statements that follow it but precede subsequent MODEL statements.

Currently, only linear equality restrictions are permitted in PROC PANEL. Tests and restriction expressions can only be composed of algebraic operations that involve the addition symbol (+), subtraction symbol (–), and multiplication symbol (*).

The RESTRICT statement accepts labels that are produced in the printed output. A RESTRICT statement can be labeled in two ways. A RESTRICT statement can be preceded by a label followed by a colon. This is illustrated in **rest1** in the example below. Alternatively, the keyword RESTRICT can be followed by a quoted string as illustrated by **"rest2"** in the example.

The following statements illustrate the use of the RESTRICT statement:

```
proc panel;
   model y = x1 x2 x3;
   restrict x1 = 0, x2 * .5 + 2 * x3= 0;
   rest1: restrict  x2 = 0, x3 = 0;
   restrict "rest2" intercept=1;
run;
```

Note that a restrict statement cannot include a division sign in its formulation.

TEST Statement

TEST < *"string"*> *equation* < *,equation2. . .* >< */ options*> **;**

The TEST statement performs Wald, Lagrange multiplier and likelihood ratio tests of linear hypotheses about the regression parameters in the preceding MODEL statement. Like RESTRICT statements, TEST statements before the first MODEL statement are automatically associated with the first MODEL statement, in addition to any TEST statements that follow it but precede subsequent MODEL statements. 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 options can be specified on the TEST statement after the slash (/):

ALL
specifies Wald, Lagrange multiplier and likelihood ratio tests.

WALD
specifies the Wald test.

LM
specifies the Lagrange multiplier test.

LR
specifies the likelihood ratio test.

The Wald test is performed by default.

The following statements illustrate the use of the TEST statement:

```
proc panel;
   id csid tsid;
   model y = x1 x2 x3;
   test x1 = 0, x2 * .5 + 2 * x3 = 0;
   test_int: test intercept = 0, x3 = 0;
run;
```

The first test investigates the joint hypothesis that

$$\beta_1 = 0$$

and

$$.5\beta_2 + 2\beta_3 = 0$$

Currently, only linear equality restrictions and tests are permitted in PROC PANEL. Tests and restriction expressions can be composed only of algebraic operations that involve the addition symbol (+), subtraction symbol (–), and multiplication symbol (*).

The TEST statement accepts labels that are produced in the printed output. A TEST statement can be labeled in two ways. A TEST statement can be preceded by a label followed by a colon. Alternatively, the keyword TEST can be followed by a quoted string. If both are present, PROC PANEL uses the quoted string. If you do not supply a label, PROC PANEL automatically labels the test. If both a TEST and a RESTRICT statement are specified, the test is run with the restrictions applied.

For the DaSilva, Hausman-Taylor, and Amemiya-MaCurdy methods, only the Wald test is available.

Details: PANEL Procedure

Specifying the Input Data

The PANEL procedure is similar to other regression procedures in SAS. Suppose 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 input data set used by PROC PANEL must be sorted by cross section and by time within each cross section. Therefore, the first step in PROC PANEL is to make sure that the input data set is sorted. The following statements sort the data set A appropriately:

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

The next step is to invoke the PANEL procedure and specify the cross section and time series variables in an ID statement. The following statements show the correct syntax:

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

Alternatively, PROC PANEL has the capability to read "flat" data. Say that you are using the data set A, which has observations on states. Specifically, the data are composed of observations on Y, X1, and X2. Unlike the previous case, the data are not recorded with a PROC PANEL structure. Instead, you have all of a state's information on a single row. You have variables to denote the name of the state (such as state). The time observations for the Y variable are recorded horizontally. So the variable Y_1 is the first period's time observation, and the variable Y_10 is the tenth period's observation for some state. The same holds for the other variables. You have the variables X1_1 to X1_10, X2_1 to X2_10, and X3_1 to X3_10 for others. With such data, PROC PANEL could be called by using the following syntax:

```
proc panel data=a;
   flatdata indid = state base = (Y X1 X2) tsname = t;
   id state t;
   model Y = X1 X2;
run;
```

See "FLATDATA Statement" on page 1802 and Example 26.7 for more information about the use of the FLATDATA statement.

Specifying the Regression Model

The MODEL statement in PROC PANEL 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, as shown in the following statements:

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

The major advantage of using PROC PANEL 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 PANEL procedure are FIXONE, FIXONETIME, FIXTWO, FDONE, FDONETIME, FDTWO, RANONE, RANTWO, PARKS, DASILVA, GMM1, GMM2, and ITGMM (iterated GMM). See the following sections for more information about these methods and the error structures they assume. The following statements fit a Fuller-Battese one-way random-effects model:

```
proc panel data=a;
   id state date;
   model y = x1 x2 / ranone vcomp=fb;
run;
```

You can specify more than one error structure option in the MODEL statement; the analysis is repeated using each specified method. You can use any number of MODEL statements to estimate different regression models or estimate the same model by using different options. See Example 26.1 for more information.

To aid in model specification within this class of models, PROC PANEL provides two specification test statistics. The first is an F statistic that tests the null hypothesis that the fixed-effects parameters are all 0. The second is a Hausman m statistic that provides information about the appropriateness of the random-effects specification. The m statistic 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. However, 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 0. Rejection of the null hypothesis might suggest that the fixed-effects model is more appropriate.

The PANEL procedure also provides the Buse R-square measure. 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.

Unbalanced Data

For fixed-effects models, random-effects models, between estimators, and dynamic panel estimators, the PANEL procedure can process data with different numbers of time series observations across different cross sections. While the Hausman-Taylor estimator supports unbalanced data, the closely-related Amemiya-MaCurdy estimator requires the data be balanced. The Parks and Da Silva methods cannot be used with 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.

Missing Values

Any observation in the input data set with a missing value for one or more of the regressors is ignored by PROC PANEL and is not used in the model fit.

If there are observations in the input data set with missing dependent variable values but with nonmissing regressors, PROC PANEL can compute predicted values and store them in an output data set by using the OUTPUT statement. Note that the presence of such observations with missing dependent variable values does not affect the model fit because these observations are excluded from the calculation.

If either some regressors or the dependent variable values are missing, the model is estimated as unbalanced where the number of time series observations across different cross sections does not have to be equal. The Parks and Da Silva methods cannot be used with unbalanced data.

Computational Resources

The more parameters there are to be estimated, the more memory and time are required to estimate the model. Also affecting these resources are the estimation method chosen and the method to calculate variance components. If the model has p parameters including the intercept, there are at least $p + p(p+1)/2$ numbers being held in the memory.

If the Arellano and Bond GMM approach is used, the amount of memory grows proportionately to the number of instruments in the INSTRUMENT statement. If the ITGMM (iterated GMM) option is selected, the computation time also depends on the convergence criteria selected and the maximum number of iterations allowed.

Restricted Estimates

When estimating a linear model with restrictions, the error degrees of freedom increases by the number of restrictions. PROC PANEL produces the Lagrange multiplier associated with each restriction.

Suppose that you are interested in linear regression in which there are r restrictions. A linear restriction implies the following set of equations that relate the regression coefficients:

$$\begin{aligned} R_{1,1}\beta_1 + R_{1,2}\beta_2 + \cdots + R_{1,p}\beta_p &= q_1 \\ R_{2,1}\beta_1 + R_{2,2}\beta_2 + \cdots + R_{2,p}\beta_p &= q_2 \\ &\vdots \\ R_{r,1}\beta_1 + R_{r,2}\beta_2 + \cdots + R_{r,p}\beta_p &= q_r \end{aligned}$$

To economize on notation, the restriction structure can be represented using matrix notation as $\mathbf{R}\boldsymbol{\beta} = \mathbf{q}$. Let $\hat{\boldsymbol{\beta}}$ be the unrestricted estimator of $\boldsymbol{\beta}$, and $\mathbf{X}$ be the corresponding set of regressors.

The regression problem then becomes the minimization of the sum of square errors subject to the restrictions. In matrix terms, the Lagrangian for this problem is then:

$$L = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) + 2\boldsymbol{\lambda}(\mathbf{R}\boldsymbol{\beta} - \mathbf{q}) = \mathbf{y}'\mathbf{y} - 2\mathbf{y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} + 2\boldsymbol{\lambda}'\mathbf{R}\boldsymbol{\beta} - 2\boldsymbol{\lambda}'\mathbf{q}$$

Minimizing the Lagrangian with respect to $\boldsymbol{\beta}$ followed by some rearranging yields the following expression for the restricted estimator $\boldsymbol{\beta}$:

$$\boldsymbol{\beta}^* = \hat{\boldsymbol{\beta}} - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\boldsymbol{\lambda}$$

From the restriction, we know that $\mathbf{R}\boldsymbol{\beta}^* = \mathbf{q}$. This allows us then to solve for the Lagrange multipliers $\boldsymbol{\lambda}$. $\boldsymbol{\lambda}$ is then:

$$\boldsymbol{\lambda}_* = \left[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right]^{-1}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{q})$$

The standard errors of the Lagrange multipliers are calculated as:

$$\mathrm{Var}(\boldsymbol{\lambda}_*) = \left[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right]^{-1}\mathbf{R}\mathrm{Var}(\hat{\boldsymbol{\beta}})\mathbf{R}'\left[\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'\right]^{-1}$$

A significant Lagrange multiplier implies that you can reject the null hypothesis that the restriction is not binding.

Note that in the special case of the fixed-effects models, the NOINT option and RESTRICT INTERCEPT=0 option give different estimates. This is not an error; it reflects two perspectives on the same issue. In the FIXONE case, the intercept is the last cross section's fixed effect (or the last time affecting the case of FIXONETIME). Specifying the NOINT option removes the intercept, but allows the last effect in. The NOINT command simply reclassifies the effects. The dummy variables become true cross section effects. If you specify the NOINT option with the FIXTWO option, the restriction is imposed that the last time effect is zero. A RESTRICT INTERCEPT=0 statement suppresses the estimation of the last effect in the FIXONE and FIXONETIME case. A RESTRICT INTERCEPT=0 has similar effects on the FIXTWO estimator. In general, restricting the intercept to zero is not recommended because OLS loses its unbiased nature. Restrictions are specified by the RESTRICT statement discussed in the section entitled "RESTRICT Statement" on page 1824.

Notation

The following notation represents the usual panel structure, with the specification of u_{it} dependent on the particular model:

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

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 $\mathbf{X}$ matrix without the intercept. All other notation is specific to each section.

One-Way Fixed-Effects Model

The specification for the one-way fixed-effects model is

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

where the γ_is are nonrandom parameters to be estimated.

Let $\mathbf{Q_0} = \text{diag}(\mathbf{E}_{T_i})$, with $\bar{\mathbf{J}}_{T_i} = \mathbf{J}_{T_i}/T_i$ and $\mathbf{E}_{T_i} = \mathbf{I}_{T_i} - \bar{\mathbf{J}}_{T_i}$, where $\mathbf{J}_{T_i}$ is a matrix of T_i ones.

The matrix $\mathbf{Q_0}$ represents the within transformation. In the one-way model, the within transformation is the conversion of the raw data to deviations from a cross section's mean. The vector $\tilde{\mathbf{x}}_{it}$ is a row of the general matrix $\mathbf{X}_s$, where the subscripted *s* implies that the constant (column of ones) is missing.

Let $\tilde{\mathbf{X}}_s = \mathbf{Q_0}\mathbf{X}_s$ and $\tilde{\mathbf{y}} = \mathbf{Q_0}\mathbf{y}$. The estimator of the slope coefficients is given by

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

Once the slope estimates are in hand, the estimation of an intercept or the cross-sectional fixed effects is handled as follows. First, you obtain the cross-sectional effects:

$$\gamma_i = \bar{y}_{i\cdot} - \tilde{\beta}_s\bar{x}_{i\cdot} \quad \text{for} \quad i = 1 \ldots N$$

If the NOINT option is specified, then the dummy variables' coefficients are set equal to the fixed effects. If an intercept is desired, then the *i*th dummy variable is obtained from the following expression:

$$D_i = \gamma_i - \gamma_N \quad \text{for} \quad i = 1 \ldots N-1$$

The intercept is the *N*th fixed effect γ_N.

The within model sum of squared errors is

$$\text{SSE} = \sum_{i=1}^{N}\sum_{t=1}^{T_i}(y_{it} - \gamma_i - \mathbf{X}_{sit}\tilde{\beta}_s)^2$$

The estimated error variance can be written

$$\hat{\sigma}_\epsilon^2 = \text{SSE}/(M - N - (K-1))$$

Alternatively, an equivalent way to express the error variance is

$$\hat{\sigma}_\epsilon^2 = \tilde{\mathbf{u}}'\mathbf{Q_0}\tilde{\mathbf{u}}/(M - N - (K-1))$$

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 drawback is that the formula changes (but the results do not) with the inclusion of a constant.

The variance covariance matrix of $\tilde{\beta}_s$ is given by

$$\text{Var}\left[\tilde{\beta}_s\right] = \hat{\sigma}_\epsilon^2(\tilde{\mathbf{X}}_s'\tilde{\mathbf{X}}_s)^{-1}$$

The covariance of the dummy variables and the dummy variables with the $\tilde{\beta}_s$ is dependent on whether the intercept is included in the model.

- *no intercept:*

$$\text{Var}\left[\gamma_i\right] = \text{Var}\left[D_i\right] = \frac{\hat{\sigma}_\epsilon^2}{T_i} + \bar{\mathbf{x}}_{i\cdot}'\text{Var}\left[\tilde{\beta}_s\right]\bar{\mathbf{x}}_{i\cdot}$$

$$\text{Cov}\left[\gamma_i, \gamma_j\right] = \text{Cov}\left[D_i, D_j\right] = \bar{\mathbf{x}}_{i\cdot}'\text{Var}\left[\tilde{\beta}_s\right]\bar{\mathbf{x}}_{j\cdot}$$

$$\text{Cov}\left[\gamma_i, \tilde{\beta}_s\right] = \text{Cov}\left[D_i, \tilde{\beta}_s\right] = -\bar{\mathbf{x}}_{i\cdot}'\text{Var}\left[\tilde{\beta}_s\right]$$

- *intercept:*

$$\text{Var}\left[D_i\right] = \frac{\hat{\sigma}_\epsilon^2}{T_i} + \frac{\hat{\sigma}_\epsilon^2}{T_N} + (\bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{N\cdot})' \text{Var}\left[\tilde{\beta}_s\right] (\bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{N\cdot})$$

$$\text{Cov}\left[D_i, D_j\right] = \frac{\hat{\sigma}_\epsilon^2}{T_N} + (\bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{N\cdot})' \text{Var}\left[\tilde{\beta}_s\right] (\bar{\mathbf{x}}_{j\cdot} - \bar{\mathbf{x}}_{N\cdot})$$

$$\text{Var}\left[\text{Intercept}\right] = \text{Var}\left[\gamma_N\right] = \frac{\hat{\sigma}_\epsilon^2}{T_N} + \bar{\mathbf{x}}_{N\cdot}' \text{Var}\left[\tilde{\beta}_s\right] \bar{\mathbf{x}}_{N\cdot}$$

$$\text{Cov}\left[D_i, \tilde{\beta}_s\right] = -(\bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{N\cdot})' \text{Var}\left[\tilde{\beta}_s\right]$$

$$\text{Cov}\left[\text{Intercept}, D_i\right] = -\frac{\hat{\sigma}_\epsilon^2}{T_i} + \bar{\mathbf{x}}_{N\cdot}' \text{Var}\left[\tilde{\beta}_s\right] (\bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{N\cdot})$$

$$\text{Cov}\left[\text{Intercept}, \tilde{\beta}_s\right] = -\bar{\mathbf{x}}_{N\cdot}' \text{Var}\left[\tilde{\beta}_s\right]$$

Alternatively, the model option FIXONETIME estimates a one-way model where the heterogeneity comes from time effects. This option is analogous to re-sorting the data by time and then by cross section and running a FIXONE model. The advantage of using the FIXONETIME option is that sorting is avoided and the model remains labeled correctly.

Two-Way Fixed-Effects Model

The specification for the two-way fixed-effects model is

$$u_{it} = \gamma_i + \alpha_t + \epsilon_{it}$$

where the γ_is and α_ts are nonrandom parameters to be estimated.

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

Balanced Panels

If the data are balanced (for example, all cross sections have T observations), then you can write the following:

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

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

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 two-way fixed-effects model is simply a regression of $\tilde{y}_{it}$ on $\tilde{\mathbf{x}}_{it}$. Therefore, the two-way β is given by

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

The calculations of cross section dummy variables, time dummy variables, and intercepts follow in a fashion similar to that used in the one-way model.

First, you obtain the net cross-sectional and time effects. Denote the cross-sectional effects by γ and the time effects by α. These effects are calculated from the following relations:

$$\hat{\gamma}_i = \left(\bar{y}_{i\cdot} - \bar{\bar{y}}\right) - \tilde{\beta}_s\left(\bar{x}_{i\cdot} - \bar{\bar{x}}\right)$$

$$\hat{\alpha}_t = \left(\bar{y}_{\cdot t} - \bar{\bar{y}}\right) - \tilde{\beta}_s\left(\bar{x}_{\cdot t} - \bar{\bar{x}}\right)$$

Denote the cross-sectional dummy variables and time dummy variables with the superscript C and T. Under the NOINT option, the following equations give the dummy variables:

$$D_i^C = \hat{\gamma}_i + \hat{\alpha}_T$$

$$D_t^T = \hat{\alpha}_t - \hat{\alpha}_T$$

When an intercept is specified, the equations for dummy variables and intercept are

$$D_i^C = \hat{\gamma}_i - \hat{\gamma}_N$$

$$D_t^T = \hat{\alpha}_t - \hat{\alpha}_T$$

$$\text{Intercept} = \hat{\gamma}_N + \hat{\alpha}_T$$

The sum of squared errors is

$$\text{SSE} = \sum_{i=1}^{N}\sum_{t=1}^{T_i}(y_{it} - \gamma_i - \alpha_t - \mathbf{X}_s\tilde{\beta}_s)^2$$

The estimated error variance is

$$\hat{\sigma}_\epsilon^2 = \text{SSE}/(M - N - T - (K-1))$$

With or without a constant, the variance covariance matrix of $\tilde{\beta}_s$ is given by

$$\text{Var}\left[\tilde{\beta}_s\right] = \hat{\sigma}_\epsilon^2(\tilde{\mathbf{X}}_s'\tilde{\mathbf{X}}_s)^{-1}$$

Variance Covariance of Dummy Variables with No Intercept

The variances and covariances of the dummy variables are given with the NOINT specification as follows:

$$\begin{aligned}\operatorname{Var}\left(D_i^C\right) &= \hat{\sigma}_\epsilon^2\left(\frac{1}{T}+\frac{1}{N}-\frac{1}{NT}\right)\\ &+ \left(\bar{\mathbf{x}}_{i\cdot}+\bar{\mathbf{x}}_{\cdot T}-\bar{\bar{\mathbf{x}}}\right)'\operatorname{Var}\left[\tilde{\beta}_s\right]\left(\bar{\mathbf{x}}_{i\cdot}+\bar{\mathbf{x}}_{\cdot T}-\bar{\bar{\mathbf{x}}}\right)\end{aligned}$$

$$\operatorname{Var}\left(D_t^T\right)=\frac{2\hat{\sigma}_\epsilon^2}{N}+\left(\bar{\mathbf{x}}_{\cdot t}-\bar{\mathbf{x}}_{\cdot T}\right)'\operatorname{Var}\left[\tilde{\beta}_s\right]\left(\bar{\mathbf{x}}_{\cdot t}-\bar{\mathbf{x}}_{\cdot T}\right)$$

$$\begin{aligned}\operatorname{Cov}\left(D_i^C,D_j^C\right) &= \hat{\sigma}_\epsilon^2\left(\frac{1}{N}-\frac{1}{NT}\right)\\ &+ \left(\bar{\mathbf{x}}_{i\cdot}+\bar{\mathbf{x}}_{\cdot t}-\bar{\bar{\mathbf{x}}}\right)'\operatorname{Var}\left[\tilde{\beta}_s\right]\left(\bar{\mathbf{x}}_{j\cdot}+\bar{\mathbf{x}}_{\cdot t}-\bar{\bar{\mathbf{x}}}\right)\end{aligned}$$

$$\operatorname{Cov}\left(D_t^T,D_u^T\right)=\frac{\hat{\sigma}_\epsilon^2}{N}+\left(\bar{\mathbf{x}}_{\cdot t}-\bar{\mathbf{x}}_{\cdot T}\right)'\operatorname{Var}\left[\tilde{\beta}_s\right]\left(\bar{\mathbf{x}}_{\cdot u}-\bar{\mathbf{x}}_{\cdot T}\right)$$

$$\operatorname{Cov}\left(D_i^C,D_t^T\right)=-\frac{\hat{\sigma}_\epsilon^2}{N}+\left(\bar{\mathbf{x}}_{i\cdot}+\bar{\mathbf{x}}_{\cdot t}-\bar{\bar{\mathbf{x}}}\right)'\operatorname{Var}\left[\tilde{\beta}_s\right]\left(\bar{\mathbf{x}}_{\cdot t}-\bar{\mathbf{x}}_{\cdot T}\right)$$

$$\operatorname{Cov}\left(D_i^C,\beta\right)=-\left(\bar{\mathbf{x}}_{i\cdot}+\bar{\mathbf{x}}_{\cdot t}-\bar{\bar{\mathbf{x}}}\right)'\operatorname{Var}\left[\tilde{\beta}_s\right]$$

$$\operatorname{Cov}\left(D_i^T,\beta\right)=-\left(\bar{\mathbf{x}}_{\cdot t}-\bar{\mathbf{x}}_{\cdot T}\right)'\operatorname{Var}\left[\tilde{\beta}_s\right]$$

Variance Covariance of Dummy Variables with Intercept

The variances and covariances of the dummy variables are given when the intercept is included as follows:

$$\begin{aligned}
\mathrm{Var}\left(D_i^C\right) &= \frac{2\hat{\sigma}_\epsilon^2}{T} + \left(\bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{N\cdot}\right)' \mathrm{Var}\left[\tilde{\beta}_s\right] \left(\bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{N\cdot}\right) \\
\mathrm{Var}\left(D_t^T\right) &= \frac{2\hat{\sigma}_\epsilon^2}{N} + \left(\bar{\mathbf{x}}_{\cdot t} - \bar{\mathbf{x}}_{\cdot T}\right)' \mathrm{Var}\left[\tilde{\beta}_s\right] \left(\bar{\mathbf{x}}_{\cdot t} - \bar{\mathbf{x}}_{\cdot T}\right) \\
\mathrm{Var}\left(\mathrm{Intercept}\right) &= \hat{\sigma}_\epsilon^2 \left(\frac{1}{T} + \frac{1}{N} - \frac{1}{NT}\right) + \left(\bar{\mathbf{x}}_{N\cdot} + \bar{\mathbf{x}}_{\cdot T} - \bar{\bar{\mathbf{x}}}\right)' \mathrm{Var}\left[\tilde{\beta}_s\right] \left(\bar{\mathbf{x}}_{N\cdot} + \bar{\mathbf{x}}_{\cdot T} - \bar{\bar{\mathbf{x}}}\right) \\
\mathrm{Cov}\left(D_i^C, D_j^C\right) &= \frac{\hat{\sigma}_\epsilon^2}{T} + \left(\bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{N\cdot}\right)' \mathrm{Var}\left[\tilde{\beta}_s\right] \left(\bar{\mathbf{x}}_{j\cdot} - \bar{\mathbf{x}}_{N\cdot}\right) \\
\mathrm{Cov}\left(D_t^T, D_u^T\right) &= \frac{\hat{\sigma}_\epsilon^2}{N} + \left(\bar{\mathbf{x}}_{\cdot t} - \bar{\mathbf{x}}_{\cdot T}\right)' \mathrm{Var}\left[\tilde{\beta}_s\right] \left(\bar{\mathbf{x}}_{\cdot u} - \bar{\mathbf{x}}_{\cdot T}\right) \\
\mathrm{Cov}\left(D_i^C, D_u^T\right) &= \left(\bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{N\cdot}\right)' \mathrm{Var}\left[\tilde{\beta}_s\right] \left(\bar{\mathbf{x}}_{\cdot u} - \bar{\mathbf{x}}_{\cdot T}\right) \\
\mathrm{Cov}\left(D_i^C, \mathrm{Intercept}\right) &= -\left(\frac{\hat{\sigma}_\epsilon^2}{T}\right) + \left(\bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{N\cdot}\right)' \mathrm{Var}\left(\tilde{\beta}_s\right) \left(\bar{\mathbf{x}}_{N\cdot} + \bar{\mathbf{x}}_{\cdot T} - \bar{\bar{\mathbf{x}}}\right) \\
\mathrm{Cov}\left(D_t^T, \mathrm{Intercept}\right) &= -\left(\frac{\hat{\sigma}_\epsilon^2}{N}\right) + \left(\bar{\mathbf{x}}_{\cdot t} - \bar{\mathbf{x}}_{\cdot T}\right)' \mathrm{Var}\left[\tilde{\beta}_s\right] \left(\bar{\mathbf{x}}_{N\cdot} + \bar{\mathbf{x}}_{\cdot T} - \bar{\bar{\mathbf{x}}}\right) \\
\mathrm{Cov}\left(D_i^C, \tilde{\beta}\right) &= -\left(\bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{N\cdot}\right)' \mathrm{Var}\left[\tilde{\beta}_s\right] \\
\mathrm{Cov}\left(D_t^T, \tilde{\beta}\right) &= -\left(\bar{\mathbf{x}}_{\cdot t} - \bar{\mathbf{x}}_{\cdot T}\right)' \mathrm{Var}\left[\tilde{\beta}_s\right] \\
\mathrm{Cov}\left(\mathrm{Intercept}, \tilde{\beta}\right) &= -\left(\bar{\mathbf{x}}_{N\cdot} + \bar{\mathbf{x}}_{\cdot T} - \bar{\bar{\mathbf{x}}}\right)' \mathrm{Var}\left[\tilde{\beta}_s\right]
\end{aligned}$$

Unbalanced Panels

Let $\mathbf{X}_*$ and $\mathbf{y}_*$ be the independent and dependent variables arranged by time and by cross section within each time period. (Note that the input data set used by the PANEL procedure must be sorted by cross section and then by time within each cross section.) 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 that correspond 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 $\mathbf{Z}_2 = \mathrm{diag}(\mathbf{D}_1\mathbf{j}_N, \mathbf{D}_2\mathbf{j}_N, \ldots, \mathbf{D}_T\mathbf{j}_N)$. The matrix $\mathbf{Z}$ gives the dummy variable structure for the two-way model.

Let

$$\begin{aligned}
\mathbf{\Delta}_N &= \mathbf{Z}_1'\mathbf{Z}_1 \\
\mathbf{\Delta}_T &= \mathbf{Z}_2'\mathbf{Z}_2 \\
\mathbf{A} &= \mathbf{Z}_2'\mathbf{Z}_1 \\
\bar{\mathbf{Z}} &= \mathbf{Z}_2 - \mathbf{Z}_1\Delta_N^{-1}\mathbf{A}' \\
\mathbf{Q} &= \Delta_T - \mathbf{A}\Delta_N^{-1}\mathbf{A}' \\
\mathbf{P} &= (\mathbf{I}_M - \mathbf{Z}_1\Delta_N^{-1}\mathbf{Z}_1') - \bar{\mathbf{Z}}\mathbf{Q}^{-1}\bar{\mathbf{Z}}'
\end{aligned}$$

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 actual implementation is quite different from the theory. The PANEL procedure transforms all series using the $\mathbf{P}$ matrix.

$$\bar{\mathbf{v}} = \mathbf{P}\mathbf{v}$$

The variable being transformed is v, which could be $\mathbf{y}$ or any column of $\mathbf{X}$. After the data are properly transformed, OLS is run on the resulting series.

Given $\tilde{\beta}_s$, the next step is estimating the cross-sectional and time effects. Given that $\boldsymbol{\gamma}$ is the column vector of cross-sectional effects and $\boldsymbol{\alpha}$ is the column vector of time effects,

$$\tilde{\boldsymbol{\alpha}} = \mathbf{Q}^{-1}\bar{\mathbf{Z}}^{'}y - \mathbf{Q}^{-1}\bar{\mathbf{Z}}^{'}\mathbf{X}_s\tilde{\beta}_s$$

$$\tilde{\boldsymbol{\gamma}} = (\Theta_1 + \Theta_2 - \Theta_3)y - (\Theta_1 + \Theta_2 - \Theta_3)\mathbf{X}_s\tilde{\beta}_s$$

$$\Theta_1 = \Delta_N^{-1}\mathbf{Z}_1^{'}$$

$$\Theta_2 = \Delta_N^{-1}\mathbf{A}^{'}Q^{-1}\mathbf{Z}_2^{'}$$

$$\Theta_3 = \Delta_N^{-1}\mathbf{A}^{'}Q^{-1}\mathbf{A}\Delta_N^{-1}\mathbf{Z}_1^{'}$$

Given the cross-sectional and time effects, the next step is to derive the associated dummy variables. Using the NOINT option, the following equations give the dummy variables:

$$D_i^C = \hat{\gamma}_i + \hat{\alpha}_T$$

$$D_t^T = \hat{\alpha}_t - \hat{\alpha}_T$$

When an intercept is desired, the equations for dummy variables and intercept are

$$D_i^C = \hat{\gamma}_i - \hat{\gamma}_N$$

$$D_t^T = \hat{\alpha}_t - \hat{\alpha}_T$$

$$\text{Intercept} = \hat{\gamma}_N + \hat{\alpha}_T$$

The calculation of the covariance matrix is as follows:

$$\begin{aligned}\text{Var}\,[\hat{\boldsymbol{\gamma}}] &= \hat{\sigma}_{\epsilon}^2\left(\Delta_N^{-1} - \Sigma_1 + \Sigma_2\right) \\ &+ (\Theta_1 + \Theta_2 - \Theta_3)\text{Var}\left[\tilde{\beta}_s\right](\Theta_1 + \Theta_2 - \Theta_3)^{'}\end{aligned}$$

where

$$\Sigma_1 = \Delta_N^{-1}\mathbf{A}'\mathbf{Q}^{-1}\mathbf{A}\Delta_N^{-1}\mathbf{A}'\mathbf{Q}^{-1}\mathbf{A}\Delta_N^{-1}$$

$$\Sigma_2 = \Delta_N^{-1}\mathbf{A}'\mathbf{Q}^{-1}\Delta_T\mathbf{Q}^{-1}\mathbf{A}\Delta_N$$

$$\text{Var}\left[\hat{\boldsymbol{\alpha}}\right] = \hat{\sigma}_\epsilon^2\left(\mathbf{Q}^{-1}\bar{\mathbf{Z}}'\bar{\mathbf{Z}}\mathbf{Q}^{-1}\right) + \left(\mathbf{Q}^{-1}\bar{\mathbf{Z}}'\mathbf{X}_s\right)\text{Var}\left[\tilde{\beta}_s\right]\left(\mathbf{X}_s'\bar{\mathbf{Z}}\mathbf{Q}^{-1}\right)$$

$$\begin{aligned}\text{Cov}\left[\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\gamma}}'\right] &= \hat{\sigma}_\epsilon^2\Delta_N^{-1}\left[\mathbf{A}'\mathbf{Q}^{-1}\Delta_T - \mathbf{A}'\mathbf{Q}^{-1}\mathbf{A}\Delta_N^{-1}\mathbf{A}'\right]\mathbf{Q}^{-1} \\ &+ (\Theta_1 + \Theta_2 - \Theta_3)\text{Var}\left[\tilde{\beta}_s\right]\left(\mathbf{X}_s'\bar{\mathbf{Z}}\mathbf{Q}^{-1}\right)\end{aligned}$$

$$\text{Cov}\left[\hat{\boldsymbol{\gamma}}, \tilde{\beta}\right] = (\Theta_1 + \Theta_2 - \Theta_3)\text{Var}\left[\tilde{\beta}_s\right]$$

$$\text{Cov}\left[\hat{\boldsymbol{\alpha}}, \tilde{\beta}\right] = \left(\mathbf{Q}^{-1}\bar{\mathbf{Z}}'\mathbf{X}_s\right)\text{Var}\left[\tilde{\beta}_s\right]$$

Now you work out the variance covariance estimates for the dummy variables.

Variance Covariance of Dummy Variables with No Intercept

The variances and covariances of the dummy variables are given under the NOINT selection as follows:

$$\text{Cov}\left(D_i^C, D_j^C\right) = \text{Cov}\left(\hat{\gamma}_i, \hat{\gamma}_j\right) + \text{Cov}\left(\hat{\gamma}_i, \hat{\alpha}_T\right) + \text{Cov}\left(\hat{\gamma}_j, \hat{\alpha}_T\right) + \text{Var}\left(\hat{\alpha}_T\right)$$

$$\text{Cov}\left(D_t^T, D_u^T\right) = \text{Cov}\left(\hat{\alpha}_t, \hat{\alpha}_u\right) - \text{Cov}\left(\hat{\alpha}_t, \hat{\alpha}_T\right) - \text{Cov}\left(\hat{\alpha}_u, \hat{\alpha}_T\right) + \text{Var}\left(\hat{\alpha}_T\right)$$

$$\text{Cov}\left(D_i^C, D_t^T\right) = \text{Cov}\left(\hat{\gamma}_i\, \hat{\alpha}_t\right) + \text{Cov}\left(\hat{\gamma}_i, \hat{\alpha}_T\right) - \text{Cov}\left(\hat{\gamma}_i, \hat{\alpha}_T\right) - \text{Var}\left(\hat{\alpha}_T\right)$$

$$\text{Cov}\left(D_i^C, \tilde{\beta}\right) = -\text{Cov}\left(\hat{\gamma}_i, \tilde{\beta}\right) - \text{Cov}\left(\hat{\alpha}_T, \tilde{\beta}\right)$$

$$\text{Cov}\left(D_t^T, \tilde{\beta}\right) = -\text{Cov}\left(\hat{\alpha}_t, \tilde{\beta}\right) + \text{Cov}\left(\hat{\alpha}_T, \tilde{\beta}\right)$$

Variance Covariance of Dummy Variables with Intercept

The variances and covariances of the dummy variables are given as follows when the intercept is included:

$$\text{Cov}\left(D_i^C, D_j^C\right) = \text{Cov}\left(\hat{\gamma}_i, \hat{\gamma}_j\right) - \text{Cov}\left(\hat{\gamma}_i, \hat{\gamma}_N\right) - \text{Cov}\left(\hat{\gamma}_j, \hat{\gamma}_N\right) + \text{Var}\left(\hat{\gamma}_N\right)$$

$$\text{Cov}\left(D_t^T, D_u^T\right) = \text{Cov}\left(\hat{\alpha}_t, \hat{\alpha}_u\right) - \text{Cov}\left(\hat{\alpha}_t, \hat{\alpha}_T\right) - \text{Cov}\left(\hat{\alpha}_u, \hat{\alpha}_T\right) + \text{Var}\left(\hat{\alpha}_T\right)$$

$$\text{Cov}\left(D_i^C, D_t^T\right) = \text{Cov}\left(\hat{\gamma}_i, \hat{\alpha}_t\right) - \text{Cov}\left(\hat{\gamma}_i, \hat{\alpha}_T\right) - \text{Cov}\left(\hat{\gamma}_N, \hat{\alpha}_t\right) + \text{Cov}\left(\hat{\gamma}_N, \hat{\alpha}_T\right)$$

$$\text{Cov}\left(D_i^C, \text{Intercept}\right) = \text{Cov}\left(\hat{\gamma}_i, \hat{\gamma}_N\right) + \text{Cov}\left(\hat{\gamma}_i, \hat{\alpha}_T\right) - \text{Cov}\left(\hat{\gamma}_j, \hat{\alpha}_T\right) - \text{Var}\left(\hat{\gamma}_N\right)$$

$$\text{Cov}\left(D_t^T, \text{Intercept}\right) = \text{Cov}\left(\hat{\alpha}_t, \hat{\alpha}_T\right) + \text{Cov}\left(\hat{\alpha}_t, \hat{\gamma}_N\right) - \text{Cov}\left(\hat{\alpha}_T, \hat{\alpha}_N\right) - \text{Var}\left(\hat{\alpha}_T\right)$$

$$\text{Cov}\left(D_i^C, \tilde{\beta}\right) = -\text{Cov}\left(\hat{\gamma}_i, \tilde{\beta}\right) - \text{Cov}\left(\hat{\gamma}_N, \tilde{\beta}\right)$$

$$\text{Cov}\left(D_t^T, \tilde{\beta}\right) = -\text{Cov}\left(\hat{\alpha}_t, \tilde{\beta}\right) + \text{Cov}\left(\hat{\alpha}_T, \tilde{\beta}\right)$$

$$\text{Cov}\left(\text{Intercept}, \tilde{\beta}_f\right) = -\text{Cov}\left(\hat{\alpha}_T, \tilde{\beta}\right) - \text{Cov}\left(\hat{\gamma}_N, \tilde{\beta}\right)$$

First-Differenced Methods for One-Way and Two-Way Models

The first-differenced (FD) estimator is an approach that is used to address the problem of omitted variables in econometrics and statistics by using panel data. The estimator is obtained by running a pooled OLS estimation for a regression of the differenced variables. The specification of the models, along with the estimation of the fixed effects, is the same as that described in the sections "One-Way Fixed-Effects Model" on page 1829 and "Two-Way Fixed-Effects Model" on page 1831. To eliminate the fixed effects in this model, you use first-differenced methods to difference them out instead of using the within transformation. Because the intercept is differenced out, the intercept cannot be estimated by first-differenced methods.

Let i be the cross sections and t be the time periods. The regressors and dependent variables are denoted as $\mathbf{X}_{i,t}$ and $\mathbf{y}_{i,t}$, respectively. For the models that have only cross-sectional effects, the data are transformed by first-differencing within each cross section. Therefore, the transformed variables are $\tilde{\mathbf{X}}_{i,t} = \mathbf{X}_{i,t} - \mathbf{X}_{i,t-1}$ for regressors and $\tilde{\mathbf{y}}_{i,t} = \mathbf{y}_{i,t} - \mathbf{y}_{i,t-1}$ for the dependent variable.

For models that have only time effects, the transformation is $\tilde{\mathbf{X}}_{i,t} = \mathbf{X}_{i,t} - \mathbf{X}_{i-1,t}$ for regressors and $\tilde{\mathbf{y}}_{i,t} = \mathbf{y}_{i,t} - \mathbf{y}_{i-1,t}$ for the dependent variable.

For models that have both cross-sectional effects and time effects, the transformation is $\tilde{\mathbf{X}}_{s,t} = \mathbf{X}_{s,t} - \mathbf{X}_{i-1,t} - \mathbf{X}_{i,t-1} + \mathbf{X}_{i-1,t-1}$ for regressors and $\tilde{\mathbf{y}}_{i,t} = \mathbf{y}_{i,t} - \mathbf{y}_{i-1,t} - \mathbf{y}_{i,t-1} + \mathbf{y}_{i-1,t-1}$ for the dependent variable.

The first-differenced estimator is

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

The resulting residual can be denoted as $\tilde{u} = \tilde{\mathbf{y}} - \tilde{\beta}_{fd}\tilde{\mathbf{X}}$. The number of degrees of freedom is the same as in a one-way fixed-effects model or a two-way fixed-effects model when the within transformation is used.

To calculate the predicted value, you can use the previous time period or last individual's information or both. If the model has only cross-sectional effects, the predicted value is $\hat{y}_{it} = y_{i,t-1} + \tilde{u}_{i,t}$. If the model has only time effects, the predicted value is $\hat{y}_{it} = y_{i-1,t} + \tilde{u}_{i,t}$. If the model has both cross-sectional and time effects, the predicted value is $\hat{y}_{it} = y_{i,t-1} + y_{i-1,t} - y_{i-1,t-1} + \tilde{u}_{i,t}$.

Between Estimators

The between-groups estimator is the regression of the cross section means of $\mathbf{y}$ on the cross section means of $\tilde{\mathbf{X}}_s$. In other words, you fit the following regression:

$$\bar{y}_{i\cdot} = \bar{\mathbf{x}}_{i\cdot}\beta^{BG} + \eta_i$$

The between-time-periods estimator is the regression of the time means of $\mathbf{y}$ on the time means of $\tilde{\mathbf{X}}_s$. In other words, you fit the following regression:

$$\bar{y}_{\cdot t} = \bar{\mathbf{x}}_{\cdot t}\beta^{BT} + \zeta_t$$

In either case, the error is assumed to be normally distributed with mean zero and a constant variance.

Pooled Estimator

PROC PANEL allows you to pool time series cross-sectional data and run regressions on the data. Pooling is admissible if there are no fixed effects or random effects present in the data. This feature is included to aid in analysis and comparison across model types and to give you access to HCCME standard errors and other panel diagnostics. In general, this model type should not be used with time series cross-sectional data.

One-Way Random-Effects Model

The specification for the one-way random-effects model is

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

Let $\mathbf{j}_{T_i}$ (lowercase j) be a vector of ones of dimension T_i, and $\mathbf{J}_{T_i}$ (uppercase J) be a square matrix of ones of dimension T_i. Define $\mathbf{Z_0} = \mathrm{diag}(\mathbf{j}_{T_i})$, $\mathbf{P_0} = \mathrm{diag}(\bar{\mathbf{J}}_{T_i})$, and $\mathbf{Q_0} = \mathrm{diag}(\mathbf{E}_{T_i})$, 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 the transformations $\tilde{\mathbf{X}}_s = \mathbf{Q_0}\mathbf{X}_s$ and $\tilde{\mathbf{y}} = \mathbf{Q_0}\mathbf{y}$.

In the one-way model, estimation proceeds in a two-step fashion. First, you obtain estimates of the variance of the σ_ϵ^2 and σ_v^2. There are multiple ways to derive these estimates; PROC PANEL provides four options. All four options are valid for balanced or unbalanced panels. Once these estimates are in hand, they are used to form a weighting factor θ, and estimation proceeds via OLS on partial deviations from group means.

PROC PANEL provides four options for estimating variance components, as described in what follows.

Fuller and Battese Method

The Fuller and Battese method for estimating variance components can be obtained with the option VCOMP = FB and the option RANONE. The variance components are given by the following equations (For the approach in the two-way model, see Baltagi and Chang (1994); Fuller and Battese (1974)). Let

$$R(v) = \mathbf{y}'\mathbf{Z_0}(\mathbf{Z_0}'\mathbf{Z_0})^{-1}\mathbf{Z_0}'\mathbf{y}$$

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

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

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

The estimator of the error variance is given by

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

If the NOINT option is specified, the estimator is

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

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

$$\hat{\sigma}_v^2 = \{R(v|\beta) - (N-1)\hat{\sigma}_\epsilon^2\} / \left\{M - \mathrm{tr}(\mathbf{Z_0}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z_0})\right\}$$

Note that the error variance is the variance of the residual of the within estimator.

According to Baltagi and Chang (1994), the Fuller and Battese method is appropriate to apply to both balanced and unbalanced data. The Fuller and Battese method is the default for estimation of one-way random-effects models with balanced panels. However, the Fuller and Battese method does not always obtain nonnegative estimates for the cross section (or group) variance. In the case of a negative estimate, a warning is printed and the estimate is set to zero.

Wansbeek and Kapteyn Method

The Wansbeek and Kapteyn method for estimating variance components can be obtained by setting VCOMP = WK (together with the option RANONE). The estimation of the one-way unbalanced data model is performed by using a specialization (Baltagi and Chang 1994) of the approach used by Wansbeek and Kapteyn (1989) for unbalanced two-way models. The Wansbeek and Kapteyn method is the default for unbalanced data. If just RANONE is specified, without the VCOMP= option, PROC PANEL estimates the variance component under Wansbeek and Kapteyn's method.

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 $\tilde{\mathbf{u}} = (\mathbf{I}_M - \mathbf{J}_M)\{\mathbf{y} - \mathbf{X}_s(\tilde{\mathbf{X}}'_s\tilde{\mathbf{X}}_s)^{-1}\tilde{\mathbf{X}}_s'\tilde{\mathbf{y}}\}$ 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

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

$$E(q_2) = (N - 1 + \mathrm{tr}[(\mathbf{X}_s^{'}\mathbf{Q}_0\mathbf{X}_s)^{-1}\mathbf{X}_s^{'}\mathbf{P}_0\mathbf{X}_s] - \mathrm{tr}[(\mathbf{X}_s^{'}\mathbf{Q}_0\mathbf{X}_s)^{-1}\mathbf{X}_s^{'}\bar{\mathbf{J}}_M\mathbf{X}_s])\sigma_\epsilon^2$$
$$[M - (\sum_i T_i^2/M)]\sigma_\nu^2$$

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

The estimator of the error variance is the residual variance of the within estimate. The Wansbeek and Kapteyn method can also generate negative variance components estimates.

Wallace and Hussain Method

The Wallace and Hussain method for estimating variance components can be obtained by setting VCOMP = WH (together with the option RANONE). Wallace-Hussain estimates start from OLS residuals on a data that are assumed to exhibit groupwise heteroscedasticity. As in the Wansbeek and Kapteyn method, you start with

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

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

However, instead of using the 'true' errors, you substitute the OLS residuals. You solve the system

$$E(\hat{q}_1) = E(\hat{u}'_{OLS}\mathbf{Q}_0\hat{u}_{OLS}) = \delta_{11}\hat{\sigma}_\nu^2 + \delta_{12}\hat{\sigma}_\epsilon^2$$

$$E(\hat{q}_2) = E(\hat{u}'_{OLS}\mathbf{P}_0\hat{u}_{OLS}) = \delta_{21}\hat{\sigma}_\nu^2 + \delta_{22}\hat{\sigma}_\epsilon^2$$

The constants $\delta_{11}, \delta_{12}, \delta_{21}, \delta_{22}$ are given by

$$\delta_{11} = \mathrm{tr}\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}_0\mathbf{Z}_0'\mathbf{X}\right) - \mathrm{tr}\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{P}_0\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}_0\mathbf{Z}_0'\mathbf{X}\right)$$

$$\delta_{12} = M - N - K + \mathrm{tr}\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{P}_0\mathbf{X}\right)$$

$$\delta_{21} = M - 2tr\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}_0\mathbf{Z}_0'\mathbf{X}\right) + \mathrm{tr}\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{P}_0\mathbf{X}\right)$$

$$\delta_{22} = N - \mathrm{tr}\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{P}_0\mathbf{X}\right)$$

where tr() is the trace operator on a square matrix.

Solving this system produces the variance components. This method is applicable to balanced and unbalanced panels. However, there is no guarantee of positive variance components. Any negative values are fixed equal to zero.

Nerlove Method

The Nerlove method for estimating variance components can be obtained by setting VCOMP = NL. The Nerlove method (see Baltagi 2008, p. 20) is assured to give estimates of the variance components that are always positive. Furthermore, it is simple in contrast to the previous estimators.

If γ_i is the *i*th fixed effect, Nerlove's method uses the variance of the fixed effects as the estimate of $\hat{\sigma}_\nu^2$. You have $\hat{\sigma}_\nu^2 = \sum_{i=1}^{N} \frac{(\gamma_i - \bar{\gamma})^2}{N-1}$, where $\bar{\gamma}$ is the mean fixed effect. The estimate of σ_ϵ^2 is simply the residual sum of squares of the one-way fixed-effects regression divided by the number of observations.

Transformation and Estimation

After you calculate the variance components from any method, the next task is to estimate the regression model of interest. For each individual, you form a weight (θ_i) as

$$\theta_i = 1 - \sigma_\epsilon / w_i$$

$$w_i^2 = T_i\sigma_\nu^2 + \sigma_\epsilon^2$$

where T_i is the *i*th cross section's time observations.

Taking the θ_i, you form the partial deviations,

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

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

where $\bar{y}_{i\cdot}$ and $\bar{x}_{i\cdot}$ are cross-sectional means of the dependent variable and independent variables (including the constant if any), respectively.

The random effects β is then the result of simple OLS on the transformed data.

Two-Way Random-Effects Model

The specification for the two-way random-effects model is

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

As in the one-way random-effects model, the PANEL procedure provides four options for variance component estimators. Unlike the one-way random-effects model, unbalanced panels present some special concerns.

Let $\mathbf{X}_*$ and $\mathbf{y}_*$ be the independent and dependent variables arranged by time and by cross section within each time period. (Note that the input data set used by the PANEL procedure must be sorted by cross section and then by time within each cross section.) 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 that correspond 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 $\mathbf{Z}_2 = \mathrm{diag}(\mathbf{D}_1\mathbf{j}_N, \mathbf{D}_2\mathbf{j}_N, \ldots, \mathbf{D}_T\mathbf{j}_N)$.

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

For notational ease, let

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

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

$$\bar{\Delta}_1 = \mathbf{I}_M - \mathbf{Z}_1\Delta_N^{-1}\mathbf{Z}_1'$$

$$\bar{\Delta}_2 = \mathbf{I}_M - \mathbf{Z}_2\Delta_T^{-1}\mathbf{Z}_2'$$

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

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

Fuller and Battese Method

The Fuller and Battese method for estimating variance components can be obtained by setting VCOMP = FB (with the option RANTWO). FB is the default method for a RANTWO model with balanced panel. If RANTWO is requested without specifying the VCOMP= option, PROC PANEL proceeds under the Fuller and Battese method.

Following the discussion in Baltagi, Song, and Jung (2002), the Fuller and Battese method forms the estimates as follows.

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 $\mathbf{P}$ is the Wansbeek and Kapteyn within estimator for unbalanced (or balanced) panel in a two-way setting.

The estimator of the error variance is the same as that in the Wansbeek and Kapteyn method.

Consider the expected values

$$\begin{aligned} E(q_N) &= \sigma_\epsilon^2 [M - T - K + 1] \\ &+ \sigma_\nu^2 \left[M - T - \mathrm{tr} \left(\mathbf{X}_s^{'} \bar{\Delta}_2 \mathbf{Z}_1 \mathbf{Z}_1^{'} \bar{\Delta}_2 \mathbf{X}_s \left(\mathbf{X}_s^{'} \bar{\Delta}_2 \mathbf{X}_s \right)^{-1} \right) \right] \\ E(q_T) &= \sigma_\epsilon^2 [M - N - K + 1] \\ &+ \sigma_e^2 \left[M - N - \mathrm{tr} \left(\mathbf{X}_s^{'} \bar{\Delta}_1 \mathbf{Z}_2 \mathbf{Z}_2^{'} \bar{\Delta}_1 \mathbf{X}_s \left(\mathbf{X}_s^{'} \bar{\Delta}_1 \mathbf{X}_s \right)^{-1} \right) \right] \end{aligned}$$

Just as in the one-way case, there is always the possibility that the (estimated) variance components will be negative. In such a case, the negative components are fixed to equal zero. After substituting the group sum of the within residuals for (q_N), the time sums of the within residuals for (q_T), and $\hat{\sigma}_\epsilon^2$, the two equations are solved for $\hat{\sigma}_\nu^2$ and $\hat{\sigma}_e^2$.

Wansbeek and Kapteyn Method

The Wansbeek and Kapteyn method for estimating variance components can be obtained by setting VCOMP = WK. The following methodology, outlined in Wansbeek and Kapteyn (1989) is used to handle both balanced and unbalanced data. The Wansbeek and Kapteyn method is the default for a RANTWO model with unbalanced panel. If RANTWO is requested without specifying the VCOMP= option, PROC PANEL proceeds under the Wansbeek and Kapteyn method if the panel is unbalanced.

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 computing 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

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

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

The expected values are

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

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

where

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

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

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

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

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

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

When the NOINT option is specified, the variance component equations change slightly. In particular, the following is true (Wansbeek and Kapteyn 1989):

$$E(q_N) = (T + k_N)\sigma^2 + T\sigma_v^2 + M\sigma_e^2$$

$$E(q_T) = (N + k_T)\sigma^2 + M\sigma_v^2 + N\sigma_e^2$$

Wallace and Hussain Method

The Wallace and Hussain method for estimating variance components can be obtained by setting VCOMP = WH. Wallace and Hussain's method is by far the most computationally intensive. It uses the OLS residuals to estimate the variance components. In other words, the Wallace and Hussain method assumes that the following holds:

$$q_\epsilon = \tilde{\mathbf{u}}_{OLS}^{'}\mathbf{P}\tilde{\mathbf{u}}_{OLS}$$

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

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

Taking expectations yields

$$E(q_\epsilon) = E\left(\tilde{\mathbf{u}}_{OLS}^{'}\mathbf{P}\tilde{\mathbf{u}}_{OLS}\right) = \delta_{11}\sigma_\epsilon^2 + \delta_{12}\sigma_v^2 + \delta_{13}\sigma_e^2$$

$$E(q_N) = E\left(\tilde{\mathbf{u}}_{OLS}^{'}\mathbf{Z}_2\Delta_T^{-1}\mathbf{Z}_2^{'}\tilde{\mathbf{u}}_{OLS}\right) = \delta_{21}\sigma_\epsilon^2 + \delta_{22}\sigma_v^2 + \delta_{23}\sigma_e^2$$

$$E(q_T) = E\left(\tilde{\mathbf{u}}_{OLS}^{'}\mathbf{Z}_1\Delta_N^{-1}\mathbf{Z}_1^{'}\tilde{\mathbf{u}}_{OLS}\right) = \delta_{31}\sigma_\epsilon^2 + \delta_{32}\sigma_v^2 + \delta_{33}\sigma_e^2$$

where the δ_{js} constants are defined by

$$\delta_{11} = M - N - T + 1 - \mathrm{tr}\left(\mathbf{X}^{'}\mathbf{P}\mathbf{X}\left(\mathbf{X}^{'}\mathbf{X}\right)^{-1}\right)$$

$$\delta_{12} = \mathrm{tr}\left(\mathbf{X}^{'}\mathbf{Z}_1\mathbf{Z}_1^{'}\mathbf{X}\left(\mathbf{X}^{'}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{'}\mathbf{P}\mathbf{X}\left(\mathbf{X}^{'}\mathbf{X}\right)^{-1}\right)\right)$$

$$\delta_{13} = \mathrm{tr}\left(\mathbf{X}^{'}\mathbf{Z}_2\mathbf{Z}_2^{'}\mathbf{X}\left(\mathbf{X}^{'}\mathbf{X}\right)^{-1}\left(\mathbf{X}^{'}\mathbf{P}\mathbf{X}\left(\mathbf{X}^{'}\mathbf{X}\right)^{-1}\right)\right)$$

$$\delta_{21} = T - \mathrm{tr}\left(\mathbf{X}^{'}\mathbf{Z}_2\Delta_T^{-1}\mathbf{Z}_2^{'}\mathbf{X}\left(\mathbf{X}^{'}\mathbf{X}\right)^{-1}\right)$$

$$\begin{aligned}\delta_{22} &= T - 2\mathrm{tr}\left(\mathbf{X}'\mathbf{Z}_2\Delta_T^{-1}\mathbf{Z}_2'\mathbf{Z}_1\mathbf{Z}_1'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)\\ &+ \mathrm{tr}\left(\mathbf{X}'\mathbf{Z}_2\Delta_T^{-1}\mathbf{Z}_2'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}_1\mathbf{Z}_1'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)\end{aligned}$$

$$\begin{aligned}\delta_{23} &= M - 2\mathrm{tr}\left(\mathbf{X}'\mathbf{Z}_2\mathbf{Z}_2'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)\\ &+ \mathrm{tr}\left(\mathbf{X}'\mathbf{Z}_2\Delta_T^{-1}\mathbf{Z}_2'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}_2\mathbf{Z}_2'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)\end{aligned}$$

$$\delta_{31} = N - \mathrm{tr}\left(\mathbf{X}'\mathbf{Z}_1\Delta_N^{-1}\mathbf{Z}_1'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$$

$$\begin{aligned}\delta_{32} &= M - 2\mathrm{tr}\left(\mathbf{X}'\mathbf{Z}_1\mathbf{Z}_1'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)\\ &+ \mathrm{tr}\left(\mathbf{X}'\mathbf{Z}_1\Delta_N^{-1}\mathbf{Z}_1'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}_1\mathbf{Z}_1'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)\end{aligned}$$

$$\begin{aligned}\delta_{33} &= N - 2\mathrm{tr}\left(\mathbf{X}'\mathbf{Z}_1\Delta_N^{-1}\mathbf{Z}_1'\mathbf{Z}_2\mathbf{Z}_2'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)\\ &+ \mathrm{tr}\left(\mathbf{X}'\mathbf{Z}_1\Delta_N^{-1}\mathbf{Z}_1'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Z}_2\mathbf{Z}_2'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)\end{aligned}$$

The PANEL procedure solves this system for the estimates $\hat{\sigma}_\epsilon$, $\hat{\sigma}_v$, and $\hat{\sigma}_e$. Some of the estimated variance components can be negative. Negative components are set to zero and estimation proceeds.

Nerlove Method

The Nerlove method for estimating variance components can be obtained with by setting VCOMP=NL.

The estimator of the error variance is

$$\hat{\sigma}_\epsilon^2 = \tilde{\mathbf{u}}'\mathbf{P}\tilde{\mathbf{u}}/M$$

The variance components for cross section and time effects are

$$\hat{\sigma}_v^2 = \sum_{i=1}^{N}\frac{(\gamma_i - \bar{\gamma})^2}{N-1} \text{ where } \gamma_i \text{ is the } i\text{th cross section effect}$$

and

$$\hat{\sigma}_e^2 = \sum_{i=1}^{T}\frac{(\alpha_t - \bar{\alpha})^2}{T-1} \text{ where } \alpha_i \text{ is the } t\text{th time effect}$$

Transformation and Estimation

After you calculate the estimates of the variance components, you can proceed to the final estimation. If the panel is balanced, partial mean deviations are used:

$$\tilde{y}_{it} = y_{it} - \theta_1 \bar{y}_{i\cdot} - \theta_2 \bar{y}_{\cdot t} + \theta_3 \bar{y}_{\cdot\cdot}$$

$$\tilde{x}_{it} = x_{it} - \theta_1 \bar{x}_{i\cdot} - \theta_2 \bar{x}_{\cdot t} + \theta_3 \bar{x}_{\cdot\cdot}$$

The θ estimates are obtained from

$$\theta_1 = 1 - \frac{\sigma_\epsilon}{\sqrt{T\sigma_\nu^2 + \sigma_\epsilon^2}}$$

$$\theta_2 = 1 - \frac{\sigma_\epsilon}{\sqrt{N\sigma_e^2 + \sigma_\epsilon^2}}$$

$$\theta_3 = \theta_1 + \theta_2 + \frac{\sigma_\epsilon}{\sqrt{T\sigma_\nu^2 + N\sigma_e^2 + \sigma_\epsilon^2}} - 1;$$

With these partial deviations, PROC PANEL uses OLS on the transformed series (including an intercept if you want).

The case of an unbalanced panel is somewhat more complicated. You could naively substitute the variance components in the following equation:

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

After inverting the expression for Ω, it is possible to do GLS on the data (even if the panel is unbalanced). However, the inversion of Ω is no small matter because the dimension is at least $\frac{M(M+1)}{2}$.

Wansbeek and Kapteyn show that the inverse of Ω can be written as

$$\sigma_\epsilon^2 \Omega^{-1} = \mathbf{V} - \mathbf{V}\mathbf{Z}_2 \tilde{\mathbf{P}}^{-1} \mathbf{Z}_2' \mathbf{V}$$

with the following:

$$\begin{aligned}
\mathbf{V} &= \mathbf{I}_M - \mathbf{Z}_1 \tilde{\Delta}_N^{-1} \mathbf{Z}_1' \\
\tilde{\mathbf{P}} &= \tilde{\Delta}_T - \mathbf{A} \tilde{\Delta}_N^{-1} \mathbf{A}' \\
\tilde{\Delta}_N &= \Delta_N + \left(\frac{\sigma_\epsilon^2}{\sigma_\nu^2} \right) \mathbf{I}_N \\
\tilde{\Delta}_T &= \Delta_T + \left(\frac{\sigma_\epsilon^2}{\sigma_e^2} \right) \mathbf{I}_T
\end{aligned}$$

Computationally, this is a much less intensive approach.

By using the inverse of the covariance matrix of the error, it becomes possible to complete GLS on the unbalanced panel, using the transform $\tilde{\mathbf{y}_*} = \sigma_\epsilon \Omega^{-1/2} \mathbf{y}_*$, and similarly for the regressors.

Hausman-Taylor Estimation

The Hausman and Taylor (1981) model is a hybrid that combines the consistency of a fixed-effects model with the efficiency and applicability of a random-effects model. One-way random-effects models assume exogeneity of the regressors, namely that they be independent of both the cross-sectional and observation-level errors. In cases where some regressors are correlated with the cross-sectional errors, the random effects model can be adjusted to deal with the endogeneity.

Consider the one-way model:

$$y_{it} = \mathbf{X}_{1it}\boldsymbol{\beta}_1 + \mathbf{X}_{2it}\boldsymbol{\beta}_2 + \mathbf{Z}_{1i}\boldsymbol{\gamma}_1 + \mathbf{Z}_{2i}\boldsymbol{\gamma}_2 + \nu_i + \epsilon_{it}$$

The regressors are subdivided so that the $\mathbf{X}$ variables vary within cross sections whereas the $\mathbf{Z}$ variables do not and would otherwise be dropped from a fixed-effects model. The subscript 1 denotes variables that are independent of both error terms (exogenous variables), and the subscript 2 denotes variables that are independent of the observation-level errors ϵ_{it} but correlated with cross-sectional errors ν_i (endogenous variables). The intercept term (if your model has one) is included as part of $\mathbf{Z}_1$ in what follows.

The Hausman-Taylor estimator is an instrumental variables regression on data that are weighted similarly to data for random-effects estimation. In both cases, the weights are functions of the estimated variance components.

Begin with $\mathbf{P}_0 = \mathrm{diag}(\bar{\mathbf{J}}_{T_i})$ and $\mathbf{Q}_0 = \mathrm{diag}(\mathbf{E}_{T_i})$. The mean transformation vector is $\bar{\mathbf{J}}_{T_i} = \mathbf{J}_{T_i}/T_i$ and the deviations from the mean transform is $\mathbf{E}_{T_i} = \mathbf{I}_{T_i} - \bar{\mathbf{J}}_{T_i}$, where $\mathbf{J}_{T_i}$ is a square matrix of ones of dimension T_i.

The observation-level variance is estimated from a standard fixed-effects model fit. For $\mathbf{X}_s = \{\mathbf{X}_1, \mathbf{X}_2\}$, $\tilde{\mathbf{X}}_s = \mathbf{Q}_0\mathbf{X}_s$, and $\tilde{\mathbf{y}} = \mathbf{Q}_0\mathbf{y}$, let

$$\begin{aligned}
\tilde{\boldsymbol{\beta}}_s &= \left(\tilde{\mathbf{X}}_s'\tilde{\mathbf{X}}_s\right)^{-1}\tilde{\mathbf{X}}_s'\tilde{\mathbf{y}} \\
\mathrm{SSE} &= \left(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}_s\tilde{\boldsymbol{\beta}}_s\right)'\left(\tilde{\mathbf{y}} - \tilde{\mathbf{X}}_s\tilde{\boldsymbol{\beta}}_s\right) \\
\hat{\sigma}_\epsilon^2 &= \mathrm{SSE}/(M-N)
\end{aligned}$$

To estimate the cross-sectional error variance, form the mean residuals $\mathbf{r} = \mathbf{P}_0'(\mathbf{y} - \mathbf{X}_s\tilde{\boldsymbol{\beta}}_s)$. You can use the mean residuals to obtain intermediate estimates of the coefficients for $\mathbf{Z}_1$ and $\mathbf{Z}_2$ via two-stage least squares (2SLS) estimation. At the first stage, use $\mathbf{X}_1$ and $\mathbf{Z}_1$ as instrumental variables to predict $\mathbf{Z}_2$. At the second stage, regress $\mathbf{r}$ on both $\mathbf{Z}_1$ and the predicted $\mathbf{Z}_2$ to obtain $\hat{\boldsymbol{\gamma}}_1^m$ and $\hat{\boldsymbol{\gamma}}_2^m$.

To estimate the cross-sectional variance, form $\hat{\sigma}_\nu^2 = \{R(\nu)/N - \hat{\sigma}_\epsilon^2\}/\bar{T}$, with $\bar{T} = N/(\sum_{i=1}^N T_i^{-1})$ and

$$R(\nu) = \left(\mathbf{r} - \mathbf{Z}_1\hat{\boldsymbol{\gamma}}_1^m - \mathbf{Z}_2\hat{\boldsymbol{\gamma}}_2^m\right)'\left(\mathbf{r} - \mathbf{Z}_1\hat{\boldsymbol{\gamma}}_1^m - \mathbf{Z}_2\hat{\boldsymbol{\gamma}}_2^m\right)$$

After variance-component estimation, transform the dependent variable into partial deviations: $y_{it}^* = y_{it} - \hat{\theta}_i\bar{y}_{i\cdot}$. Likewise, transform the regressors to form $\mathbf{X}_{1it}^*$, $\mathbf{X}_{2it}^*$, $\mathbf{Z}_{1i}^*$, and $\mathbf{Z}_{2i}^*$. The partial weights $\hat{\theta}_i$ are determined by $\hat{\theta}_i = 1 - \hat{\sigma}_\epsilon/\hat{w}_i$, with $\hat{w}_i^2 = T_i\hat{\sigma}_\nu^2 + \hat{\sigma}_\epsilon^2$.

Finally, you obtain the Hausman-Taylor estimates by performing 2SLS regression of y_{it}^* on $\mathbf{X}_{1it}^*$, $\mathbf{X}_{2it}^*$, $\mathbf{Z}_{1i}^*$, and $\mathbf{Z}_{2i}^*$. For the first-stage regression, use the following instruments:

- $\tilde{\mathbf{X}}_{it}$, the deviations from cross-sectional means for all time-varying variables $\mathbf{X}$, for the ith cross section during time period t
- $(1-\hat{\theta}_i)\bar{\mathbf{X}}_{1i.}$, where $\bar{\mathbf{X}}_{1i.}$ are the means of the time-varying exogenous variables for the ith cross section
- $(1-\hat{\theta}_i)\mathbf{Z}_{1i}$

Multiplication by the factor $(1-\hat{\theta}_i)$ is redundant in balanced data, but necessary in the unbalanced case to produce accurate instrumentation; see Gardner (1998).

Let k_1 equal the number of regressors in $\mathbf{X}_1$, and g_2 equal the number of regressors in $\mathbf{Z}_2$. Then the Hausman-Taylor model is identified only if $k_1 \geq g_2$; otherwise, no estimation will take place.

Hausman and Taylor (1981) describe a specification test that compares their model to fixed effects. For a null hypothesis of fixed effects, Hausman's m statistic is calculated by comparing the parameter estimates and variance matrices for both models, identically to how it is calculated for one-way random effects models; for more information, see the section "Specification Tests" on page 1870. The degrees of freedom of the test, however, are not based on matrix rank but instead are equal to $k_1 - g_2$.

Amemiya-MaCurdy Estimation

The Amemiya and MaCurdy (1986) model is similar to the Hausman-Taylor model. Following the development in the section "Hausman-Taylor Estimation" on page 1846, estimation is identical up to the final 2SLS instrumental variables regression. In addition to the set of instruments used by the Hausman-Taylor estimator, use the following:

- $\mathbf{X}_{1i1}, \mathbf{X}_{1i2}, \ldots, \mathbf{X}_{1iT}$

For each observation in the ith cross section, you use the data on the time-varying exogenous regressors for the entire cross section. Because of the structure of the added instruments, the Amemiya-MaCurdy estimator can be applied only to balanced data.

The Amemiya-MaCurdy model attempts to gain efficiency over Hausman-Taylor by adding instruments. This comes at a price of a more stringent assumption on the exogeneity of the $\mathbf{X}_1$ variables. Although the Hausman-Taylor model requires only that the cross-sectional means of $\mathbf{X}_1$ be orthogonal to ν_i, the Amemiya-MaCurdy estimation requires orthogonality at every point in time; see Baltagi (2008, sec. 7.4).

A Hausman specification test is provided to test the validity of the added assumption. Define $\boldsymbol{\alpha}' = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2', \boldsymbol{\gamma}_1', \boldsymbol{\gamma}_2')$, its Hausman-Taylor estimate as $\hat{\boldsymbol{\alpha}}_{\mathrm{HT}}$, and its Amemiya-MaCurdy estimate as $\hat{\boldsymbol{\alpha}}_{\mathrm{AM}}$. Under the null hypothesis, both estimators are consistent and $\hat{\boldsymbol{\alpha}}_{\mathrm{AM}}$ is efficient. The Hausman test statistic is then

$$m = (\hat{\boldsymbol{\alpha}}_{\mathrm{HT}} - \hat{\boldsymbol{\alpha}}_{\mathrm{AM}})' \left(\hat{\mathbf{S}}_{\mathrm{HT}} - \hat{\mathbf{S}}_{\mathrm{AM}}\right)^{-1} (\hat{\boldsymbol{\alpha}}_{\mathrm{HT}} - \hat{\boldsymbol{\alpha}}_{\mathrm{AM}})$$

where $\hat{\mathbf{S}}_{\mathrm{HT}}$ and $\hat{\mathbf{S}}_{\mathrm{AM}}$ are variance-covariance estimates of $\hat{\boldsymbol{\alpha}}_{\mathrm{HT}}$ and $\hat{\boldsymbol{\alpha}}_{\mathrm{AM}}$, respectively. Under the null hypothesis, m is distributed as χ^2 with degrees of freedom equal to the rank of $(\hat{\mathbf{S}}_{\mathrm{HT}} - \hat{\mathbf{S}}_{\mathrm{AM}})^{-1}$.

Parks Method (Autoregressive Model)

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

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

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 $\boldsymbol{\beta}$ is then estimated by generalized least squares. The first step in estimating **V** involves the use of ordinary least squares to estimate $\boldsymbol{\beta}$ and obtain the fitted residuals, as follows:

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

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

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

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

$$y_{i1}\sqrt{1-\hat{\rho}_i^2} = \sum_{k=1}^{p} X_{i1k_ek}\sqrt{1-\hat{\rho}_i^2} + u_{i1}\sqrt{1-\hat{\rho}_i^2}$$

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

which is written

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

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

The second step in estimating the covariance matrix $\mathbf{V}$ is applying ordinary least squares to the preceding transformed model, obtaining

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

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

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

where

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

Estimated generalized least squares (EGLS) then proceeds in the usual manner,

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

where $\hat{\mathbf{V}}$ is the derived consistent estimator of $\mathbf{V}$. For computational purposes, $\hat{\boldsymbol{\beta}}_P$ is obtained directly from the transformed model,

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

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 this estimator is consistent and asymptotically, normally distributed with

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

Standard Corrections

For the PARKS option, the first-order autocorrelation coefficient must be estimated for each cross section. Let ρ be the $N \times 1$ 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 PANEL, the following modification for R is made:

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

where

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

and

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

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

Da Silva Method (Variance-Component Moving Average Model)

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

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

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.

The following conditions are assumed:

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 $\mathbf{X}$ has a full column rank p.
2. $\boldsymbol{\beta}$ is a $p \times 1$ constant vector of unknown parameters.
3. $\mathbf{a}$ is a vector of uncorrelated random variables such that $E(a_i) = 0$ and $\text{var}(a_i) = \sigma_a^2$, $\sigma_a^2 > 0, i = 1, \ldots, N$.
4. $\mathbf{b}$ is a vector of uncorrelated random variables such that $E(b_t) = 0$ and $\text{var}(b_t) = \sigma_b^2$ where $\sigma_b^2 > 0$ and $t = 1, \ldots, T$.
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,

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

   where $\alpha_0, \alpha_1, \ldots, \alpha_m$ are unknown constants such that $\alpha_0 \neq 0$ and $\alpha_m \neq 0$, and $\{\epsilon_{ij}\}_{j=-\infty}^{j=\infty}$ is a white noise process for each i—that is, a sequence of uncorrelated random variables with $E(\epsilon_t) = 0$, $E(\epsilon_t^2) = \sigma_\epsilon^2$, and $\sigma_\epsilon^2 > 0$. $\{\epsilon_{ij}\}_{j=-\infty}^{j=\infty}$ for $i = 1, \ldots, N$ are mutually uncorrelated.
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 $i = 1, \ldots, N; t = 1, \ldots, T$; and $k = 1, \ldots, m$.

If assumptions 1–6 are satisfied, then

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

and

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

where Ψ_T is a $T \times T$ matrix with elements ψ_{ts},

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

where $\psi(k) = \sigma_\epsilon^2 \sum_{j=0}^{m-k} \alpha_j \alpha_{j+k}$ for $k = |t-s|$. For the definition of I_N, I_T, J_N, and J_T, see the section "Fuller and Battese Method" on page 1838.

The covariance matrix, denoted by $\mathbf{V}$, can be written in the form

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

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

Thus, the covariance matrix of the vector of observations $\mathbf{y}$ has the form

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

where

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

The estimator of $\boldsymbol{\beta}$ is a two-step GLS-type estimator—that is, GLS with the unknown covariance matrix replaced by a suitable estimator of $\mathbf{V}$. 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 $(i = 1, \ldots, n)$ are associated with a linear model $E(\mathbf{y}) = \mathbf{X}\beta$ with covariance matrix $\sum_{i=1}^{n} \nu_i V_i$ where V_i $(i = 1, \ldots, n)$ are real symmetric matrices. The method is also discussed by Seely (1970b, a); Seely and Zyskind (1971). Seely and Soong (1971) consider the MINQUE principle, using an approach along the lines of Seely (1969).

Dynamic Panel Estimators

For an example of dynamic panel estimation using the generalized method of moments (GMM) see "Example 26.6: Dynamic Panel Estimation of Cigarette Sales Data" on page 1914.

Consider the case of the following general model:

$$y_{it} = \Sigma_{l=1}^{maxlag} \phi_l y_{i(t-l)} + \Sigma_{k=1}^{K} \beta_k x_{itk} + \gamma_i + \alpha_t + \epsilon_{it}$$

The x variables can include ones that are correlated or uncorrelated to the individual effects, predetermined, or strictly exogenous. The variable x_{it}^p is defined as predetermined in the sense that $E\left(x_{it}^p \epsilon_{is}\right) \neq 0$ for $s < t$

and zero otherwise. The variable x^e_{it} is defined as strictly exogenous if $E\left(x^e_{it}\epsilon_{is}\right) = 0$ for all s and t. The γ_i and α_t are cross-sectional and time series fixed effects, respectively. Arellano and Bond (1991) show that it is possible to define conditions that should result in a consistent estimator.

Consider the simple case of an autoregression in a panel setting (with only individual effects):

$$y_{it} = \phi y_{i(t-1)} + \gamma_i + \epsilon_{it}$$

Differencing the preceding relationship results in:

$$\Delta y_{it} = \phi \Delta y_{i(t-1)} + \nu_{it}$$

where $\nu_{it} = \epsilon_{it} - \epsilon_{it-1}$.

Obviously, y_{it} is not exogenous. However, Arellano and Bond (1991) show that it is still useful as an instrument, if properly lagged. This instrument is required with the option DEPVAR(LEVEL).

For $t = 2$ (assuming the first observation corresponds to time period 1) you have,

$$\Delta y_{i2} = \phi \Delta y_{i1} + \nu_{i2}$$

Using y_{i1} as an instrument is not a good idea since $\mathrm{Cov}\left(\epsilon_{i1}, \nu_{i2}\right) \neq 0$. Therefore, since it is not possible to form a moment restriction, you discard this observation.

For $t = 3$ you have,

$$\Delta y_{i3} = \phi \Delta y_{i2} + \nu_{i3}$$

Clearly, you have every reason to suspect that $\mathrm{Cov}\left(\epsilon_{i1}, \nu_{i3}\right) = 0$. This condition forms one restriction.

For $t = 4$, both $\mathrm{Cov}\left(\epsilon_{i1}, \nu_{i4}\right) = 0$ and $\mathrm{Cov}\left(\epsilon_{i2}, \nu_{i4}\right) = 0$ must hold.

Proceeding in that fashion, you have the following matrix of instruments,

$$\mathbf{Z}_i = \begin{pmatrix} y_{i1} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & y_{i1} & y_{i2} & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & y_{i1} & y_{i2} & y_{i3} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & & & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & y_{i1} & \cdots & y_{i(T-2)} \end{pmatrix}$$

Using the instrument matrix, you form the weighting matrix $\mathbf{A_N}$ as

$$\mathbf{A_N} = \left(\frac{1}{N}\sum_i^N \mathbf{Z}_i'\mathbf{H}_i\mathbf{Z}_i\right)^{-1}$$

The initial weighting matrix is

$$\mathbf{H}_i = \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & & & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 \end{pmatrix}$$

Note that the maximum size of the $\mathbf{H}_i$ matrix is T–2. The origins of the initial weighting matrix are the expected error covariances. Notice that on the diagonals,

$$E\left(v_{it}v_{it}\right) = E\left(\epsilon_{it}^2 - 2\epsilon_{it}\epsilon_{i(t-1)} + \epsilon_{i(t-1)}^2\right) = 2\sigma_\epsilon^2$$

and off diagonals,

$$E\left(v_{it}v_{i(t-1)}\right) = E\left(\epsilon_{it}\epsilon_{i(t-1)} - \epsilon_{it}\epsilon_{i(t-2)} - \epsilon_{i(t-1)}\epsilon_{i(t-1)} + \epsilon_{i(t-1)}\epsilon_{i(t-2)}\right) = -\sigma_\epsilon^2$$

If you let the vector of lagged differences (in the series y_{it}) be denoted as Δy_{i-} and the dependent variable as Δy_i, then the optimal GMM estimator is

$$\phi = \left[\left(\sum_i \Delta y_{i-}^{'}\mathbf{Z}_i\right)\mathbf{A_N}\left(\sum_i \mathbf{Z}_i^{'}\Delta y_{i-}\right)\right]^{-1}\left(\sum_i \Delta y_{i-}^{'}\mathbf{Z}_i\right)\mathbf{A_N}\left(\sum_i \mathbf{Z}_i^{'}\Delta y_i\right)$$

Using the estimate, $\hat{\phi}$, you can obtain estimates of the errors, $\hat{\epsilon}$, or the differences, $\hat{v}$. From the errors, the variance is calculated as,

$$\sigma^2 = \frac{\hat{\epsilon}^{'}\hat{\epsilon}}{M-1}$$

where $M = \sum_{i=1}^{N} T_i$ is the total number of observations. With differenced equations, since we lose the first two observations, $M = \sum_{i=1}^{N}\left(T_i - 2\right)$.

Furthermore, you can calculate the variance of the parameter as,

$$\sigma^2\left[\left(\Sigma_i \Delta y_{i-}^{'}\mathbf{Z}_i\right)\mathbf{A_N}\left(\sum_i \mathbf{Z}_i^{'}\Delta y_{i-}\right)\right]^{-1}$$

Alternatively, you can view the initial estimate of the ϕ as a first step. That is, by using $\hat{\phi}$, you can improve the estimate of the weight matrix, $\mathbf{A_N}$.

Instead of imposing the structure of the weighting, you form the $\mathbf{H}_i$ matrix through the following:

$$\mathbf{H}_i = \hat{\boldsymbol{\nu}}_i\hat{\boldsymbol{\nu}}_i^{'}$$

You then complete the calculation as previously shown. The PROC PANEL option GMM2 specifies this estimation.

The case of multiple right-hand-side variables illustrates more clearly the power of Arellano and Bond (1991); Arellano and Bover (1995).

Considering the general case you have:

$$y_{it} = \sum_{l=1}^{maxlag}\phi_l y_{i(t-l)} + \boldsymbol{\beta}\mathbf{X}_i + \gamma_i + \alpha_t + \epsilon_{it}$$

It is clear that lags of the dependent variable are both not exogenous and correlated to the fixed effects. However, the independent variables can fall into one of several categories. An independent variable can be

correlated[1] and exogenous, uncorrelated and exogenous, correlated and predetermined, and uncorrelated and predetermined. The category in which an independent variable is found influences when or whether it becomes a suitable instrument. Note, however, that neither PROC PANEL nor Arellano and Bond require that a regressor be an instrument or that an instrument be a regressor.

First, suppose that the variables are all correlated with the individual effects γ_i. Consider the question of exogenous or predetermined. An exogenous variable is not correlated with the error term $\epsilon_{it} - \epsilon_{i,t-1}$ in the differenced equations. Therefore, all observations (on the exogenous variable) become valid instruments at all time periods. If the model has only one instrument and it happens to be exogenous, then the optimal instrument matrix looks like,

$$\mathbf{Z}_i = \begin{pmatrix} x_{i1} \cdots x_{iT} & 0 & 0 & 0 & 0 \\ 0 & x_{i1} \cdots x_{iT} & 0 & 0 & 0 \\ 0 & 0 & x_{i1} \cdots x_{iT} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & x_{i1} \cdots x_{iT} \end{pmatrix}$$

The situation for the predetermined variables becomes a little more difficult. A predetermined variable is one whose future realizations can be correlated to current shocks in the dependent variable. With such an understanding, it is admissible to allow all current and lagged realizations as instruments. In other words you have,

$$\mathbf{Z}_i = \begin{pmatrix} x_{i1} & 0 & 0 & 0 & 0 \\ 0 & x_{i1}\, x_{i2} & 0 & 0 & 0 \\ 0 & 0 & x_{i1} \cdots x_{i3} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & x_{i1} \cdots x_{i(T-1)} \end{pmatrix}$$

When the data contain a mix of endogenous, exogenous, and predetermined variables, the instrument matrix is formed by combining the three. For example, the third observation would have one observation on the dependent variable as an instrument, three observations on the predetermined variables as instruments, and all observations on the exogenous variables.

Now consider some variables, denoted as x_{1it}, that are not correlated with the individual effects γ_i. There is yet another set of moment restrictions that can be used. An uncorrelated variable means that the variable's level is not affected by the individual specific effect. You write the preceding general model as

$$y_{it} = \sum_{l=1}^{maxlag} \phi_l y_{i(t-l)} + \Sigma_{k=1}^{K} \beta_k x_{itk} + \alpha_t + \mu_{it}$$

where $\mu_{it} = \gamma_i + \epsilon_{it}$.

Because the variables are uncorrelated with γ_i and thus uncorrelated with the error term μ_{it} in the level equations, you can use the difference and level equations to perform a system estimation. That is, the uncorrelated variables imply moment restrictions on the level equations. Given the previously used restrictions for the equations in first differences, there are T extra restrictions. For predetermined variables, Arellano

[1] In this section, "correlated" means correlated with the individual effects and "uncorrelated" means uncorrelated with the individual effects.

and Bond (1991) use the extra restrictions $E\left(\mu_{i2}x^{p}_{1i1}\right) = 0$ and $E\left(\mu_{it}x^{p}_{1it}\right) = 0$ for $t = 2, \ldots, T$. The instrument matrix becomes

$$\mathbf{Z}^{*}_{i} = \begin{pmatrix} \mathbf{Z}_i & 0 & 0 & 0 & \cdots & 0 \\ 0 & x^{p}_{1i1} & x^{p}_{1i2} & 0 & \cdots & 0 \\ 0 & 0 & 0 & x^{p}_{1i3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & x^{p}_{1iT} \end{pmatrix}$$

For exogenous variables x^{e}_{1it} Arellano and Bond (1991) use $E\left(T^{-1}\sum_{s=1}^{T}\mu_{is}x^{e}_{1it}\right) = 0$. PROC PANEL uses the same ones as the predetermined variables—that is, $E\left(\mu_{i2}x^{e}_{1i1}\right) = 0$ and $E\left(\mu_{it}x^{e}_{1it}\right) = 0$ for $t = 2, \ldots, T$. If you denote the new instrument matrix by using the full complement of instruments available by an asterisk and if both x^{p} and x^{e} are uncorrelated, then you have

$$\mathbf{Z}^{*}_{i} = \begin{pmatrix} \mathbf{Z}_i & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & x^{p}_{i1} & x^{e}_{i1} & x^{p}_{i2} & x^{e}_{i2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & x^{p}_{i3} & x^{e}_{i3} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & x^{p}_{iT} & x^{e}_{iT} \end{pmatrix}$$

When the lagged dependent variable is included as the explanatory variable (as in the dynamic panel data models), Blundell and Bond (1998) suggest the system GMM to use $T - 2$ extra-moment restrictions, which use the lagged differences as the instruments for the level:

$$E\left(\mu_{it}\Delta y_{i,t-1}\right) = 0 \qquad \text{for } t = 3, \ldots, T$$

This additional set of moment conditions are required by DEPVAR(DIFF) option. The corresponding instrument matrix is

$$\mathbf{Z}^{y}_{li} = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & \Delta y_{i2} & 0 & \cdots & 0 \\ 0 & 0 & \Delta y_{i3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \Delta y_{i(T-1)} \end{pmatrix}$$

Blundell and Bond (1998) argue that the system GMM that uses these extra conditions significantly increases the efficiency of the estimator, especially under strong serial correlation in the dependent variables.[2]

Except for those GMM-type instruments, PROC PANEL can also handle standard instruments by using the lists that you specify in the LEVELEQ= and DIFFEQ= options. Denote l_{it} and d_{it} as the standard instruments that are specified for the level equation and differenced equation, respectively. The additional moment restrictions are $E\left(\mu_{it}l_{it}\right) = 0$ for $t = 1, \ldots, T$ for level equations and $E\left(\Delta\epsilon_{it}d_{it}\right) = 0$ for $t = 2, \ldots, T$ for differenced equations. The instrument matrix for the level and differenced equations are $\mathbf{Z}_{li}$

[2]This happens when $\phi \to 1$ or as $\sigma^2_{\gamma}/\sigma^2_{\epsilon} \to \infty$. In this case, the lagged dependent variables $y_{i(t-l)}$ become weak instruments for the differenced variables Δy_{it}.

and $\mathbf{Z}_{di}$, respectively, as follows:

$$\mathbf{Z}_{li} = \begin{pmatrix} l_{i1} & 0 & 0 & 0 & 0 \\ 0 & l_{i2} & 0 & 0 & 0 \\ 0 & 0 & l_{i3} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & l_{iT} \end{pmatrix}$$

$$\mathbf{Z}_{di} = \begin{pmatrix} d_{i1} & 0 & 0 & 0 & 0 \\ 0 & d_{i2} & 0 & 0 & 0 \\ 0 & 0 & d_{i3} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & d_{iT} \end{pmatrix}$$

To put the differenced and level equations together, for the system GMM estimator, the instrument matrix can be constructed as

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{Z}_{di} & 0 & 0 & 0 & 0 \\ 0 & \mathbf{Z}_{li}^{e} & \mathbf{Z}_{li}^{p} & \mathbf{Z}_{li} & \mathbf{Z}_{li}^{y} \end{pmatrix}$$

where $\mathbf{Z}_{li}^{e}$ and $\mathbf{Z}_{li}^{p}$ correspond to the exogenous and predetermined uncorrelated variables, respectively.

The formation of the initial weighting matrix becomes somewhat problematic. If you denote the new weighting matrix with an asterisk, then you can write

$$\mathbf{A}_{\mathbf{N}}^{*} = \left(\frac{1}{N} \sum_{i}^{N} \mathbf{Z}_i^{*'} \mathbf{H}_i^{*} \mathbf{Z}_i^{*} \right)^{-1}$$

where

$$\mathbf{H}_i^{*} = \begin{pmatrix} \mathbf{H}_i & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

To finish, you write out the two equations (or two stages) that are estimated,

$$\Delta y_{it} = \boldsymbol{\beta}^{*} \Delta \mathbf{S}_{it} + \alpha_t - \alpha_{t-1} + \nu_{it}; \quad y_{it} = \boldsymbol{\beta}^{*} \mathbf{S}_{it} + \gamma_i + \alpha_t + \epsilon_{it}$$

where $\mathbf{S}_{it}$ is the matrix of all explanatory variables—lagged endogenous, exogenous, and predetermined.

Let $\mathbf{y}_{it}^{*}$ be given by

$$\mathbf{y}_{it}^{*} = \begin{pmatrix} \Delta y_{it} \\ y_{it} \end{pmatrix} \quad \boldsymbol{\beta}^{*} = \begin{pmatrix} \boldsymbol{\phi} & \boldsymbol{\beta} \end{pmatrix} \quad \mathbf{S}_{it}^{*} = \begin{pmatrix} \Delta \mathbf{S}_{it} \\ \mathbf{S}_{it} \end{pmatrix} \quad \mathbf{e}_i^{*} = \begin{pmatrix} \boldsymbol{\nu}_i \\ \boldsymbol{\mu}_i = \boldsymbol{\epsilon}_i + \gamma_i \end{pmatrix}$$

Using the preceding information, you can get the one-step GMM estimator,

$$\hat{\boldsymbol{\beta}}_{\mathbf{1}}^{*} = \left[\left(\sum_i \mathbf{S}_i^{*'} \mathbf{Z}_i^{*} \right) \mathbf{A}_{\mathbf{N}}^{*} \left(\sum_i \mathbf{Z}_i^{*'} \mathbf{S}_i^{*} \right) \right]^{-1} \left(\sum_i \mathbf{S}_i^{*'} \mathbf{Z}_i^{*} \right) \mathbf{A}_{\mathbf{N}}^{*} \left(\sum_i \mathbf{Z}_i^{*'} \mathbf{y}_i^{*} \right)$$

If the GMM2 or ITGMM option is not specified in the MODEL statement, estimation terminates here. If it terminates, you can obtain the following information.

Variance of the error term comes from the second-stage (level) equations—that is,

$$\sigma^2 = \frac{\hat{\boldsymbol{\mu}}'\hat{\boldsymbol{\mu}}}{M-p} = \frac{\left(y_{it} - \hat{\boldsymbol{\beta}}_1^* \mathbf{S}_{it}\right)'\left(y_{it} - \hat{\boldsymbol{\beta}}_1^* \mathbf{S}_{it}\right)}{M-p}$$

where p is the number of regressors and M is the number of observations as defined before.

The variance covariance matrix can be obtained from

$$\left[\left(\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\sum_i \mathbf{Z}_i^{*'}\mathbf{S}_i^*\right)\right]^{-1}\sigma^2$$

Alternatively, you can obtain a robust estimate of the variance covariance matrix by specifying the ROBUST option in the MODEL statement. Without further reestimation of the model, the $\mathbf{H}_i^*$ matrix is recalculated as

$$\mathbf{H}_{i,2}^* = \begin{pmatrix} \hat{\boldsymbol{\nu}}_i\hat{\boldsymbol{\nu}}_i' & 0 \\ 0 & \hat{\boldsymbol{\mu}}_i\hat{\boldsymbol{\mu}}_i' \end{pmatrix}$$

And the weighting matrix becomes

$$\mathbf{A}_\mathbf{N}^*\left(\hat{\boldsymbol{\beta}}_1^*\right) = \left(\frac{1}{N}\sum_i^N \mathbf{Z}_i^{*'}\mathbf{H}_{i,2}^*\mathbf{Z}_i^*\right)^{-1}$$

Using the preceding information, you construct the robust covariance matrix from the following.

Let $\mathbf{G}$ denote a temporary matrix,

$$\mathbf{G} = \left[\left(\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\sum_i \mathbf{Z}_i^{*'}\mathbf{S}_i^*\right)\right]^{-1}\left(\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*$$

The robust covariance estimate of $\hat{\boldsymbol{\beta}}_1^*$ is

$$\mathbf{V}^r\left(\hat{\boldsymbol{\beta}}_1^*\right) = \mathbf{G}\mathbf{A}_\mathbf{N}^{*-1}\left(\hat{\boldsymbol{\beta}}_1^*\right)\mathbf{G}'$$

Alternatively, you can use the new weighting matrix to form an updated estimate of the regression parameters, as requested by the GMM2 option in the MODEL statement. In short,

$$\hat{\boldsymbol{\beta}}_2^* = \left[\left(\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\hat{\boldsymbol{\beta}}_1^*\right)\left(\sum_i \mathbf{Z}_i^{*'}\mathbf{S}_i^*\right)\right]^{-1}\left(\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\hat{\boldsymbol{\beta}}_1^*\right)\left(\sum_i \mathbf{Z}_i^{*'}\mathbf{y}_i^*\right)$$

The covariance estimate of the two-step $\hat{\boldsymbol{\beta}}_2^*$ becomes

$$V\left(\hat{\boldsymbol{\beta}}_2^*\right) = \left[\left(\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\hat{\boldsymbol{\beta}}_1^*\right)\left(\sum_i \mathbf{Z}_i^{*'}\mathbf{S}_i^*\right)\right]^{-1}$$

Similarly, you construct the robust covariance matrix from the following.

Let $\mathbf{G}_2$ denote a temporary matrix,

$$\mathbf{G}_2 = \left[\left(\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\hat{\boldsymbol{\beta}}_1^*\right)\left(\sum_i \mathbf{Z}_i^{*'}\mathbf{S}_i^*\right)\right]^{-1}\left(\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\hat{\boldsymbol{\beta}}_1^*\right)$$

The robust covariance estimate of $\hat{\boldsymbol{\beta}}_2^*$ is

$$\mathbf{V}^r\left(\hat{\boldsymbol{\beta}}_2^*\right) = \mathbf{G}_2\mathbf{A}_\mathbf{N}^{*-1}\left(\hat{\boldsymbol{\beta}}_2^*\right)\mathbf{G}_2'$$

According to Arellano and Bond (1991), Blundell and Bond (1998), and many others, two-step standard errors are unreliable. Therefore, researchers often base inference on two-step parameter estimates and one-step standard errors. Windmeijer (2005) derives a small-sample bias-corrected variance that uses the first-order Taylor series approximation of the two-step GMM estimator $\hat{\boldsymbol{\beta}}_2^*$ around the true value $\boldsymbol{\beta}^*$ as a function of the one-step GMM estimator $\hat{\boldsymbol{\beta}}_1^*$,

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_2^* - \boldsymbol{\beta}^* &= \left[\left(\textstyle\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\hat{\boldsymbol{\beta}}_1^*\right)\left(\textstyle\sum_i \mathbf{Z}_i^{*'}\mathbf{S}_i^*\right)\right]^{-1}\left(\textstyle\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\hat{\boldsymbol{\beta}}_1^*\right)\left(\textstyle\sum_i \mathbf{Z}_i^{*'}\mathbf{e}_i^*\right) \\
&= \left[\left(\textstyle\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\boldsymbol{\beta}^*\right)\left(\textstyle\sum_i \mathbf{Z}_i^{*'}\mathbf{S}_i^*\right)\right]^{-1}\left(\textstyle\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\boldsymbol{\beta}^*\right)\left(\textstyle\sum_i \mathbf{Z}_i^{*'}\mathbf{e}_i^*\right) \\
&\quad + D_{\boldsymbol{\beta}^*,\mathbf{A}_\mathbf{N}^*(\boldsymbol{\beta}^*)}\left(\hat{\boldsymbol{\beta}}_1^* - \boldsymbol{\beta}^*\right) + O_p\left(N^{-1}\right)
\end{aligned}$$

where $D_{\boldsymbol{\beta}^*,\mathbf{A}_\mathbf{N}^*(\boldsymbol{\beta}^*)}$ is the first derivative of $\hat{\boldsymbol{\beta}}_2^* - \boldsymbol{\beta}^*$ with regard to $\boldsymbol{\beta}'$ evaluated at the true value $\boldsymbol{\beta}^*$. The kth column of D is

$$\begin{aligned}
&\{D_{\boldsymbol{\beta}^*,\mathbf{A}_\mathbf{N}^*(\boldsymbol{\beta}^*)}\}_k = \\
&\left[\left(\textstyle\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\boldsymbol{\beta}^*\right)\left(\textstyle\sum_i \mathbf{Z}_i^{*'}\mathbf{S}_i^*\right)\right]^{-1}\left(\textstyle\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\boldsymbol{\beta}^*\right)\frac{\partial\mathbf{A}_\mathbf{N}^{*-1}(\boldsymbol{\beta})}{\partial\beta_k}|_{\boldsymbol{\beta}^*}\mathbf{A}_\mathbf{N}^*\left(\boldsymbol{\beta}^*\right)\left(\textstyle\sum_i \mathbf{Z}_i^{*'}\mathbf{S}_i^*\right) \\
&\times\left[\left(\textstyle\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\boldsymbol{\beta}^*\right)\left(\textstyle\sum_i \mathbf{Z}_i^{*'}\mathbf{S}_i^*\right)\right]^{-1}\left(\textstyle\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\boldsymbol{\beta}^*\right)\left(\textstyle\sum_i \mathbf{Z}_i^{*'}\mathbf{e}_i^*\right) \\
&-\left[\left(\textstyle\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\boldsymbol{\beta}^*\right)\left(\textstyle\sum_i \mathbf{Z}_i^{*'}\mathbf{S}_i^*\right)\right]^{-1}\left(\textstyle\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\boldsymbol{\beta}^*\right)\frac{\partial\mathbf{A}_\mathbf{N}^{*-1}(\boldsymbol{\beta})}{\partial\beta_k}|_{\boldsymbol{\beta}^*}\mathbf{A}_\mathbf{N}^*\left(\boldsymbol{\beta}^*\right)\left(\textstyle\sum_i \mathbf{Z}_i^{*'}\mathbf{e}_i^*\right)
\end{aligned}$$

Because $\boldsymbol{\beta}^*$, $\mathbf{A}_\mathbf{N}^*\left(\boldsymbol{\beta}^*\right)$, and $\frac{\partial\mathbf{A}_\mathbf{N}^{*-1}(\boldsymbol{\beta})}{\partial\beta_k}|_{\boldsymbol{\beta}^*}$ are not feasible, you can replace them with their estimators, $\hat{\boldsymbol{\beta}}_2^*$, $\mathbf{A}_\mathbf{N}^*\left(\hat{\boldsymbol{\beta}}_1^*\right)$, and $\frac{\partial\mathbf{A}_\mathbf{N}^{*-1}(\boldsymbol{\beta})}{\partial\beta_k}|_{\hat{\boldsymbol{\beta}}_1^*}$, respectively. Denote $\hat{\mathbf{e}}_{i,2}^*$ as the second-stage error term by

$$\left[\left(\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\hat{\boldsymbol{\beta}}_1^*\right)\left(\sum_i \mathbf{Z}_i^{*'}\mathbf{S}_i^*\right)\right]^{-1}\left(\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\hat{\boldsymbol{\beta}}_1^*\right)\left(\sum_i \mathbf{Z}_i^{*'}\hat{\mathbf{e}}_{i,2}^*\right) = 0$$

and

$$\frac{\partial\mathbf{A}_\mathbf{N}^{*-1}\left(\boldsymbol{\beta}\right)}{\partial\beta_k}|_{\boldsymbol{\beta}^*} = -\frac{1}{N}\sum_i \mathbf{Z}_i^{*'}\begin{pmatrix}\Delta\mathbf{S}_{i,k}\boldsymbol{\nu}_i' + \boldsymbol{\nu}_i\Delta\mathbf{S}_{i,k}' & 0 \\ 0 & \mathbf{S}_{i,k}\boldsymbol{\mu}_i' + \boldsymbol{\mu}_i\mathbf{S}_{i,k}'\end{pmatrix}\mathbf{Z}_i^*$$

The first part vanishes and leaves

$$\begin{aligned}
\{D_{\hat{\boldsymbol{\beta}}_2^*,\mathbf{A}_\mathbf{N}^*\left(\hat{\boldsymbol{\beta}}_1^*\right)}\}_k &= \frac{1}{N}\left[\left(\textstyle\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\hat{\boldsymbol{\beta}}_1^*\right)\left(\textstyle\sum_i \mathbf{Z}_i^{*'}\mathbf{S}_i^*\right)\right]^{-1}\left(\textstyle\sum_i \mathbf{S}_i^{*'}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\hat{\boldsymbol{\beta}}_1^*\right) \\
&\quad\left(\textstyle\sum_i \mathbf{Z}_i^{*'}\begin{pmatrix}\Delta\mathbf{S}_{i,k}\hat{\boldsymbol{\nu}}_{i,1}' + \hat{\boldsymbol{\nu}}_{i,1}\Delta\mathbf{S}_{i,k}' & 0 \\ 0 & \mathbf{S}_{i,k}\hat{\boldsymbol{\mu}}_{i,1}' + \hat{\boldsymbol{\mu}}_{i,1}\mathbf{S}_{i,k}'\end{pmatrix}\mathbf{Z}_i^*\right)\mathbf{A}_\mathbf{N}^*\left(\hat{\boldsymbol{\beta}}_1^*\right)\left(\textstyle\sum_i \mathbf{Z}_i^{*'}\hat{\mathbf{e}}_{i,2}^*\right)
\end{aligned}$$

Plugging these into the Taylor expansion series yields

$$V^c\left(\hat{\boldsymbol{\beta}}_2^*\right) = V\left(\hat{\boldsymbol{\beta}}_2^*\right) + D_{\hat{\boldsymbol{\beta}}_2^*,\mathbf{A}_\mathbf{N}^*\left(\hat{\boldsymbol{\beta}}_1^*\right)} V\left(\hat{\boldsymbol{\beta}}_2^*\right) + V\left(\hat{\boldsymbol{\beta}}_2^*\right) D'_{\hat{\boldsymbol{\beta}}_2^*,\mathbf{A}_\mathbf{N}^*\left(\hat{\boldsymbol{\beta}}_1^*\right)} + D_{\hat{\boldsymbol{\beta}}_2^*,\mathbf{A}_\mathbf{N}^*\left(\hat{\boldsymbol{\beta}}_1^*\right)} \mathbf{V}^r\left(\hat{\boldsymbol{\beta}}_1^*\right) D'_{\hat{\boldsymbol{\beta}}_2^*,\mathbf{A}_\mathbf{N}^*\left(\hat{\boldsymbol{\beta}}_1^*\right)}$$

As a final note, it possible to iterate more than twice by specifying the ITGMM option. At each iteration, the parameter estimates and its variance-covariance matrix (standard or robust) can be constructed as the one-step and/or two-step GMM estimators. Such a multiple iteration should result in a more stable estimate of the covariance estimate. PROC PANEL allows two convergence criteria. Convergence can occur in the parameter estimates or in the weighting matrices. Let $\mathbf{A}^*_{\mathbf{N,k+1}}$ denote the robust covariance matrix from iteration k, which is used as the weighting matrix in iteration $k+1$. Iterate until

$$\max_{i,j \le \dim\left(\mathbf{A}^*_{\mathbf{N,k}}\right)} \frac{\left|\mathbf{A}^*_{\mathbf{N,k+1}}(i,j) - \mathbf{A}^*_{\mathbf{N,k}}(i,j)\right|}{\left|\mathbf{A}^*_{\mathbf{N,k}}(i,j)\right|} \le \text{ATOL}$$

or

$$\max_{i \le \dim\left(\boldsymbol{\beta}_k^*\right)} \frac{\left|\boldsymbol{\beta}^*_{k+1}(i) - \boldsymbol{\beta}^*_k(i)\right|}{\left|\boldsymbol{\beta}^*_k(i)\right|} \le \text{BTOL}$$

where ATOL is the tolerance for convergence in the weighting matrix and BTOL is the tolerance for convergence in the parameter estimate matrix. The default convergence criteria is BTOL = 1E–8 for PROC PANEL.

Specification Testing for Dynamic Panel

Specification tests under GMM in PROC PANEL generally follow Arellano and Bond (1991). The first test available is a Sargan/Hansen test of over-identification. The test for a one-step estimation is constructed as

$$\left(\sum_i \eta_i' \mathbf{Z}_\mathbf{i}^*\right) \mathbf{A}_\mathbf{N}^* \left(\sum_i \mathbf{Z}_\mathbf{i}^{*'} \eta_i\right) \sigma^2$$

where η_i is the stacked error term (of the differenced equation and level equation).

When the robust weighting matrix is used, the test statistic is computed as

$$\left(\sum_i \eta_i' \mathbf{Z}_\mathbf{i}^*\right) \mathbf{A}_{\mathbf{N},\mathbf{2}}^* \left(\sum_i \mathbf{Z}_\mathbf{i}^{*'} \eta_i\right)$$

This definition of the Sargan test is used for all iterated estimations. The Sargan test is distributed as a χ^2 with degrees of freedom equal to the number of moment conditions minus the number of parameters.

In addition to the Sargan test, PROC PANEL tests for autocorrelation in the residuals. These tests are distributed as standard normal. PROC PANEL tests the hypothesis that the autocorrelation of the lth lag is significant.

Define $\boldsymbol{\omega}_l$ as the lag of the differenced error, with zero padding for the missing values generated. Symbolically,

$$\boldsymbol{\omega}_{l,i} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \boldsymbol{v}_{i,2} \\ \vdots \\ \boldsymbol{v}_{i,T-1-l} \end{pmatrix}$$

You define the constant k_0 as

$$k_0\,(l) = \sum_i \boldsymbol{\omega}'_{l,i}\boldsymbol{v}_i$$

You next define the constant k_1 as

$$k_1\,(l) = \sum_i \boldsymbol{\omega}'_{l,i}\mathbf{H}_i\boldsymbol{\omega}_{l,i}$$

Note that the choice of $\mathbf{H}_i$ is dependent on the stage of estimation. If the estimation is first stage, then you would use the matrix with twos along the main diagonal, and minus ones along the primary subdiagonals. In a robust estimation or multi-step estimation, this matrix would be formed from the outer product of the residuals (from the previous step).

Define the constant k_2 as

$$k_2\,(l) = -2\left(\sum_i \boldsymbol{\omega}'_{l,i}\Delta\mathbf{S}_i\right)\mathbf{G}\left(\sum_i \Delta\mathbf{S}'_i\mathbf{Z}_i\right)\mathbf{A}_{\mathbf{N},\mathbf{k}}\left(\sum_i \mathbf{Z}'_i\mathbf{H}_i\boldsymbol{\omega}_{l,i}\right)$$

The matrix $\mathbf{G}$ is defined as

$$\mathbf{G} = \left[\left(\sum_i \Delta\mathbf{S}_i^{*\prime}\mathbf{Z}_i^*\right)\mathbf{A}^*_{\mathbf{N},\mathbf{k}}\left(\sum_i \mathbf{Z}_i^{*\prime}\Delta\mathbf{S}_i^*\right)\right]^{-1}$$

The constant k_3 is defined as

$$k_3\,(l) = \left(\sum_i \boldsymbol{\omega}'_{l,i}\Delta\mathbf{S}_i\right)V\left(\boldsymbol{\beta}^*\right)\left(\sum_i \Delta\mathbf{S}'_i\boldsymbol{\omega}_{l,i}\right)$$

Using the four quantities, the test for autoregressive structure in the differenced residual is

$$m\,(l) = \frac{k_0\,(l)}{\sqrt{k_1\,(l) + k_2\,(l) + k_3\,(l)}}$$

The m statistic is distributed as a normal random variable with mean zero and standard deviation of one.

Instrument Choice

Arellano and Bond's technique is a very useful method for dealing with any autoregressive characteristics in the data. However, there is one caveat to consider. Too many instruments bias the estimator to the within estimate. Furthermore, many instruments make this technique not scalable. The weighting matrix becomes

very large, so every operation that involves it becomes more computationally intensive. The PANEL procedure enables you to specify a bandwidth for instrument selection. For example, specifying MAXBAND=10 means that there will be at most ten time observations for each variable that enter as instruments. The default is to follow the Arellano-Bond methodology.

In specifying a maximum bandwidth, you can also specify the selection of the time observations. There are three possibilities: leading, trailing (default), and centered. The exact consequence of choosing any of those possibilities depends on the variable type (correlated, exogenous, or predetermined) and the time period of the current observation.

If the MAXBAND option is specified, then the following is true under any selection criterion (let t be the time subscript for the current observation). The first observation for the endogenous variable (as instrument) is $\max(t - \mathrm{MAXBAND}, 1)$ and the last instrument is $t - 2$. The first observation for a predetermined variable is $\max(t - \mathrm{MAXBAND}, 1)$ and the last is $t - 1$. The first and last observation for an exogenous variable is given in the following list:

- *Trailing*: If $t < \mathrm{MAXBAND}$, then the first instrument is for the first time period and the last observation is $\mathrm{MAXBAND}$. Otherwise, if $t \geq \mathrm{MAXBAND}$, then the first observation is $t - \mathrm{MAXBAND} + 1$ and the last instrument to enter is t.

- *Centered*: If $t \leq \frac{\mathrm{MAXBAND}}{2}$, then the first observation is the first time period and the last observation is $\mathrm{MAXBAND}$. If $t > T - \frac{\mathrm{MAXBAND}}{2}$, then the first instrument included is $T - \mathrm{MAXBAND} + 1$ and the last observation is T. If $\frac{\mathrm{MAXBAND}}{2} < t \leq T - \frac{\mathrm{MAXBAND}}{2}$, then the first included instrument is $t - \frac{\mathrm{MAXBAND}}{2} + 1$ and the last observation is $t + \frac{\mathrm{MAXBAND}}{2}$. If the $\mathrm{MAXBAND}$ value is an odd number, the procedure decrements by one.

- *Leading* : If $t > T - \mathrm{MAXBAND}$, then the first instrument corresponds to time period $T - \mathrm{MAXBAND} + 1$ and the last observation is T. Otherwise, if $t \leq T - \mathrm{MAXBAND}$, then the first observation is t and the last observation is $t + \mathrm{MAXBAND} + 1$.

The PANEL procedure enables you to include dummy variables to deal with the presence of time effects that are not captured by including the lagged dependent variable. The dummy variables directly affect the level equations. However, this implies that the difference of the dummy variable for time period t and $t - 1$ enters the difference equation. The first usable observation occurs at $t = 3$. If the level equation is not used in the estimation, then there is no way to identify the dummy variables. Selecting the TIME option gives the same result as that which would be obtained by creating dummy variables in the data set and using those in the regression.

The PANEL procedure gives you several options when it comes to missing values and unbalanced panel. By default, any time period for which there are missing values is skipped. The corresponding rows and columns of $\mathbf{H}$ matrices are zeroed, and the calculation is continued. Alternatively, you can elect to replace missing values and missing observations with zeros (ZERO), the overall mean of the series (OAM), the cross-sectional mean (CSM), or the time series mean (TSM).

Linear Hypothesis Testing

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

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

However, it is also possible to write the F statistic as

$$F = \frac{(\hat{\mathbf{u}}_{*}^{'}\hat{\mathbf{u}}_{*} - \hat{\mathbf{u}}^{'}\hat{\mathbf{u}})/J}{\hat{\mathbf{u}}^{'}\hat{\mathbf{u}}/(M-K)}$$

where

- $\hat{\mathbf{u}}_{*}$ is the residual vector from the restricted regression
- $\hat{\mathbf{u}}$ is the residual vector from the unrestricted regression
- J is the number of restrictions
- $(M - K)$ are the degrees of freedom, M is the number of observations, and K is the number of parameters in the model

The Wald, likelihood ratio (LR), and Lagrange multiplier (LM) tests are all related to the *F* test. You use this relationship of the *F* test to the likelihood ratio and Lagrange multiplier tests. The Wald test is calculated from its definition.

The Wald test statistic is

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

The advantage of calculating Wald in this manner is that it enables you to substitute a heteroscedasticity-corrected covariance matrix for the matrix $\mathbf{V}$. PROC PANEL makes such a substitution if you request the HCCME option in the MODEL statement.

The likelihood ratio is

$$LR = M \ln\left[1 + \frac{1}{M-K}JF\right]$$

The Lagrange multiplier test statistic is

$$LM = M\left[\frac{JF}{M - K + JF}\right]$$

where JF represents the number of restrictions multiplied by the result of the F test.

Note that only the Wald is changed when the HCCME option is selected. The LR and LM tests are unchanged.

The distribution of these test statistics is the χ^2 with degrees of freedom equal to the number of restrictions imposed (J). The three tests are asymptotically equivalent, but they have differing small sample properties. Greene (2000, p. 392) and Davidson and MacKinnon (1993, pp. 456–458) discuss the small sample properties of these statistics.

Heteroscedasticity-Corrected Covariance Matrices

The HCCME= option in the MODEL statement selects the type of heteroscedasticity-consistent covariance matrix. In the presence of heteroscedasticity, the covariance matrix has a complicated structure that can result in inefficiencies in the OLS estimates and biased estimates of the covariance matrix. The variances for cross-sectional and time dummy variables and the covariances with or between the dummy variables are not corrected for heteroscedasticity in the one-way and two-way models. Whether or not HCCME is specified, they are the same. For the two-way models, the variance and the covariances for the intercept are not corrected.[3]

Consider the simple linear model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

This discussion parallels the discussion in Davidson and MacKinnon 1993, pp. 548–562. For panel data models, we apply HCCME on the transformed data($\tilde{y}$ and $\tilde{\mathbf{X}}$). In other words, we first remove the random or fixed effects through transforming/demean the data,[4] then correct heteroscedasticity (also autocorrelation with HAC option) in the residual. The assumptions that make the linear regression best linear unbiased estimator (BLUE) are $E(\epsilon) = 0$ and $E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}^{'}) = \Omega$, where Ω has the simple structure $\sigma^2\mathbf{I}$. Heteroscedasticity results in a general covariance structure, so that it is not possible to simplify Ω. The result is the following:

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

As long as the following is true, then you are assured that the OLS estimate is consistent and unbiased:

$$plim_{n\to\infty}\left(\frac{1}{n}\mathbf{X}^{'}\boldsymbol{\epsilon}\right) = 0$$

If the regressors are nonrandom, then it is possible to write the variance of the estimated $\boldsymbol{\beta}$ as the following:

$$\mathrm{Var}\left(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}\right) = (\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{X}^{'}\Omega\mathbf{X}(\mathbf{X}^{'}\mathbf{X})^{-1}$$

The effect of structure in the covariance matrix can be ameliorated by using generalized least squares (GLS), provided that Ω^{-1} can be calculated. Using Ω^{-1}, you premultiply both sides of the regression equation,

$$L^{-1}\mathbf{y} = L^{-1}\mathbf{X}_{\boldsymbol{\epsilon}} + L^{-1}\boldsymbol{\epsilon}$$

where L denotes the Cholesky root of Ω. (that is, $\Omega = LL'$ with L lower triangular).

The resulting GLS $\boldsymbol{\beta}$ is

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

[3]The dummy variables are removed by the within transformations, so their variances and covariances cannot be calculated the same way as the other regressors. They are recovered by the formulas listed in the sections "One-Way Fixed-Effects Model" on page 1829 and "Two-Way Fixed-Effects Model" on page 1831. The formulas assume homoscedasticity, so they do not apply when HCCME is specified. Therefore, standard errors, variances, and covariances are reported only when the HCCME option is ignored. HCCME standard errors for dummy variables and intercept can be calculated by the dummy variable approach with the pooled model.

[4]For more information about transforming the data, see the sections "One-Way Fixed-Effects Model" on page 1829, "Two-Way Fixed-Effects Model" on page 1831, "One-Way Random-Effects Model" on page 1838, and "Two-Way Random-Effects Model" on page 1841.

Using the GLS $\boldsymbol{\beta}$, you can write

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\mathbf{y} \\ &= (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'(\Omega^{-1}\mathbf{X}_{\epsilon} + \Omega^{-1}\boldsymbol{\epsilon}) \\ &= \beta + (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\boldsymbol{\epsilon}\end{aligned}$$

The resulting variance expression for the GLS estimator is

$$\begin{aligned}\mathrm{Var}\left(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}\right) &= (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\Omega^{-1}\mathbf{X}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}\Omega\Omega^{-1}\mathbf{X}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\end{aligned}$$

The difference in variance between the OLS estimator and the GLS estimator can be written as

$$(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}$$

By the Gauss-Markov theorem, the difference matrix must be positive definite under most circumstances (zero if OLS and GLS are the same, when the usual classical regression assumptions are met). Thus, OLS is not efficient under a general error structure. It is crucial to realize that OLS does not produce biased results. It would suffice if you had a method for estimating a consistent covariance matrix and you used the OLS $\boldsymbol{\beta}$. Estimation of the Ω matrix is certainly not simple. The matrix is square and has M^2 elements; unless some sort of structure is assumed, it becomes an impossible problem to solve. However, the heteroscedasticity can have quite a general structure. White (1980) shows that it is not necessary to have a consistent estimate of Ω. On the contrary, it suffices to calculate an estimate of the middle expression. That is, you need an estimate of:

$$\Lambda = \mathbf{X}'\Omega\mathbf{X}$$

This matrix, Λ, is easier to estimate because its dimension is K. PROC PANEL provides the following classical HCCME estimators for Λ:

The matrix is approximated by:

- HCCME=N0:

  $$\sigma^2\mathbf{X}'\mathbf{X}$$

  This is the simple OLS estimator. If you do not specify the HCCME= option, PROC PANEL defaults to this estimator.

- HCCME=0:

  $$\sum_{i=1}^{N}\sum_{t=1}^{T_i}\hat{\epsilon}_{it}^2\mathbf{x}_{it}\mathbf{x}_{it}'$$

  where N is the number of cross sections and T_i is the number of observations in ith cross section. The $\mathbf{x}_{it}'$ is from the tth observation in the ith cross section, constituting the $(\sum_{j=1}^{i-1} T_j + t)$th row of the

matrix $\mathbf{X}$. If the CLUSTER option is specified, one extra term is added to the preceding equation so that the estimator of matrix Λ is

$$\sum_{i=1}^{N}\sum_{t=1}^{T_i}\hat{\epsilon}_{it}^2\mathbf{x}_{it}\mathbf{x}_{it}^{'}+\sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{s=1}^{t-1}\hat{\epsilon}_{it}\hat{\epsilon}_{is}\left(\mathbf{x}_{it}\mathbf{x}_{is}^{'}+\mathbf{x}_{is}\mathbf{x}_{it}^{'}\right)$$

The formula is the same as the robust variance matrix estimator in Wooldridge (2002, p. 152) and it is derived under the assumptions of section 7.3.2 of Wooldridge (2002).

- HCCME=1:

$$\frac{M}{M-K}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\hat{\epsilon}_{it}^2\mathbf{x}_{it}\mathbf{x}_{it}^{'}$$

where M is the total number of observations, $\sum_{j=1}^{N}T_j$, and K is the number of parameters. With the CLUSTER option, the estimator becomes

$$\frac{M}{M-K}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\hat{\epsilon}_{it}^2\mathbf{x}_{it}\mathbf{x}_{it}^{'}+\frac{M}{M-K}\sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{s=1}^{t-1}\hat{\epsilon}_{it}\hat{\epsilon}_{is}\left(\mathbf{x}_{it}\mathbf{x}_{is}^{'}+\mathbf{x}_{is}\mathbf{x}_{it}^{'}\right)$$

The formula is similar to the robust variance matrix estimator in Wooldridge (2002, p. 152) with the heteroskedasticity adjustment term $M/(M-K)$.

- HCCME=2:

$$\sum_{i=1}^{N}\sum_{t=1}^{T_i}\frac{\hat{\epsilon}_{it}^2}{1-\hat{h}_{it}}\mathbf{x}_{it}\mathbf{x}_{it}^{'}$$

The $\hat{h}_{it}$ term is the $(\sum_{j=1}^{i-1}T_j+t)$th diagonal element of the hat matrix. The expression for $\hat{h}_{it}$ is $\mathbf{x}_{it}^{'}(\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{x}_{it}$. The hat matrix attempts to adjust the estimates for the presence of influence or leverage points. With the CLUSTER option, the estimator becomes

$$\sum_{i=1}^{N}\sum_{t=1}^{T_i}\frac{\hat{\epsilon}_{it}^2}{1-\hat{h}_{it}}\mathbf{x}_{it}\mathbf{x}_{it}^{'}+2\sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{s=1}^{t-1}\frac{\hat{\epsilon}_{it}}{\sqrt{1-\hat{h}_{it}}}\frac{\hat{\epsilon}_{is}}{\sqrt{1-\hat{h}_{is}}}\left(\mathbf{x}_{it}\mathbf{x}_{is}^{'}+\mathbf{x}_{is}\mathbf{x}_{it}^{'}\right)$$

The formula is similar to the robust variance matrix estimator in Wooldridge (2002, p. 152) with the heteroskedasticity adjustment.

- HCCME=3:

$$\sum_{i=1}^{N}\sum_{t=1}^{T_i}\frac{\hat{\epsilon}_{it}^2}{(1-\hat{h}_{it})^2}\mathbf{x}_{it}\mathbf{x}_{it}^{'}$$

With the CLUSTER option, the estimator becomes

$$\sum_{i=1}^{N}\sum_{t=1}^{T_i}\frac{\hat{\epsilon}_{it}^2}{(1-\hat{h}_{it})^2}\mathbf{x}_{it}\mathbf{x}_{it}^{'}+2\sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{s=1}^{t-1}\frac{\hat{\epsilon}_{it}}{1-\hat{h}_{it}}\frac{\hat{\epsilon}_{is}}{1-\hat{h}_{is}}\left(\mathbf{x}_{it}\mathbf{x}_{is}^{'}+\mathbf{x}_{is}\mathbf{x}_{it}^{'}\right)$$

The formula is similar to the robust variance matrix estimator in Wooldridge (2002, p. 152) with the heteroskedasticity adjustment.

- HCCME=4: PROC PANEL includes this option for the calculation of the Arellano (1987) version of the White (1980) HCCME in the panel setting. Arellano's insight is that there are N covariance matrices in a panel, and each matrix corresponds to a cross section. Forming the White HCCME for each panel, you need to take only the average of those N estimators that yield Arellano. The details of the estimation follow. First, you arrange the data such that the first cross section occupies the first T_i observations. You treat the panels as separate regressions with the form:

  $$\mathbf{y}_i = \alpha_i \mathbf{i} + \mathbf{X}_{is}\tilde{\boldsymbol{\beta}} + \boldsymbol{\epsilon}_i$$

  The parameter estimates $\tilde{\boldsymbol{\beta}}$ and α_i are the result of least squares dummy variables (LSDV) or within estimator regressions, and $\mathbf{i}$ is a vector of ones of length T_i. The estimate of the ith cross section's $\mathbf{X}'\Omega\mathbf{X}$ matrix (where the s subscript indicates that no constant column has been suppressed to avoid confusion) is $\mathbf{X}_i'\Omega\mathbf{X}_i$. The estimate for the whole sample is:

  $$\mathbf{X}_s'\Omega\mathbf{X}_s = \sum_{i=1}^{N} \mathbf{X}_i'\Omega\mathbf{X}_i$$

  The Arellano standard error is in fact a White-Newey-West estimator with constant and equal weight on each component. In the between estimators, selecting HCCME=4 returns the HCCME=0 result since there is no 'other' variable to group by.

In their discussion, Davidson and MacKinnon (1993, p. 554) argue that HCCME=1 should always be preferred to HCCME=0. Although HCCME=3 is generally preferred to 2 and 2 is preferred to 1, the calculation of HCCME=1 is as simple as the calculation of HCCME=0. Therefore, it is clear that HCCME=1 is preferred when the calculation of the hat matrix is too tedious.

All HCCME estimators have well-defined asymptotic properties. The small sample properties are not well-known, and care must exercised when sample sizes are small.

The HCCME estimator of $\mathrm{Var}(\boldsymbol{\beta})$ is used to drive the covariance matrices for the fixed effects and the Lagrange multiplier standard errors. Robust estimates of the covariance matrix for $\boldsymbol{\beta}$ imply robust covariance matrices for all other parameters.

Heteroscedasticity- and Autocorrelation-Consistent Covariance Matrices

The HAC option in the MODEL statement selects the type of heteroscedasticity- and autocorrelation-consistent covariance matrix. As with the HCCME option, an estimator of the middle expression Λ in sandwich form is needed. With the HAC option, it is estimated as

$$\Lambda_{\mathrm{HAC}} = a\sum_{i=1}^{N}\sum_{t=1}^{T_i}\hat{\epsilon}_{it}^2\mathbf{x}_{it}\mathbf{x}_{it}' + a\sum_{i=1}^{N}\sum_{t=1}^{T_i}\sum_{s=1}^{t-1}k\left(\frac{s-t}{b}\right)\hat{\epsilon}_{it}\hat{\epsilon}_{is}\left(\mathbf{x}_{it}\mathbf{x}_{is}' + \mathbf{x}_{is}\mathbf{x}_{it}'\right)$$

where $k(.)$ is the real-valued kernel function[5], b is the bandwidth parameter, and a is the adjustment factor of small sample degrees of freedom (that is, $a = 1$ if the ADJUSTDF option is not specified and otherwise $a = NT/(NT-k)$, where k is the number of parameters including dummy variables). The types of kernel functions are listed in Table 26.3.

[5]The HCCME=0 with CLUSTER option sets $k(.) = 1$.

Table 26.3 Kernel Functions

Kernel Name	Equation
Bartlett	$k(x) = \begin{cases} 1-\lvert x\rvert & \lvert x\rvert \le 1 \\ 0 & \text{otherwise} \end{cases}$
Parzen	$k(x) = \begin{cases} 1-6x^2+6\lvert x\rvert^3 & 0 \le \lvert x\rvert \le 1/2 \\ 2(1-\lvert x\rvert)^3 & 1/2 \le \lvert x\rvert \le 1 \\ 0 & \text{otherwise} \end{cases}$
Quadratic spectral	$k(x) = \frac{25}{12\pi^2 x^2}\left(\frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5)\right)$
Truncated	$k(x) = \begin{cases} 1 & \lvert x\rvert \le 1 \\ 0 & \text{otherwise} \end{cases}$
Tukey-Hanning	$k(x) = \begin{cases} (1+\cos(\pi x))/2 & \lvert x\rvert \le 1 \\ 0 & \text{otherwise} \end{cases}$

When the BANDWIDTH=ANDREWS option is specified, the bandwidth parameter is estimated as shown in Table 26.4.

Table 26.4 Bandwidth Parameter Estimation

Kernel Name	Bandwidth Parameter
Bartlett	$b = 1.1447(\alpha(1)T)^{1/3}$
Parzen	$b = 2.6614(\alpha(2)T)^{1/5}$
Quadratic spectral	$b = 1.3221(\alpha(2)T)^{1/5}$
Truncated	$b = 0.6611(\alpha(2)T)^{1/5}$
Tukey-Hanning	$b = 1.7462(\alpha(2)T)^{1/5}$

Let $\{g_{ait}\}$ denote each series in $\{g_{it} = \hat{\epsilon}_{it}\mathbf{x}_{it}\}$, and let (ρ_a, σ_a^2) denote the corresponding estimates of the autoregressive and innovation variance parameters of the AR(1) model on $\{g_{ait}\}$, $a = 1, \ldots, k$, where the AR(1) model is parameterized as $g_{ait} = \rho g_{ait-1} + \epsilon_{ait}$ with $\mathrm{Var}(\epsilon_{ait}) = \sigma_a^2$. The terms $\alpha(1)$ and $\alpha(2)$ are estimated with the following formulas:

$$\alpha(1) = \frac{\sum_{a=1}^{k} \frac{4\rho_a^2\sigma_a^4}{(1-\rho_a)^6(1+\rho_a)^2}}{\sum_{a=1}^{k} \frac{\sigma_a^4}{(1-\rho_a)^4}}; \qquad \alpha(2) = \frac{\sum_{a=1}^{k} \frac{4\rho_a^2\sigma_a^4}{(1-\rho_a)^8}}{\sum_{a=1}^{k} \frac{\sigma_a^4}{(1-\rho_a)^4}}$$

When you specify BANDWIDTH=NEWEYWEST94, according to Newey and West (1994) the bandwidth parameter is estimated as shown in Table 26.5.

Table 26.5 Bandwidth Parameter Estimation

Kernel Name	Bandwidth Parameter
Bartlett	$b = 1.1447(\{s_1/s_0\}^2 T)^{1/3}$
Parzen	$b = 2.6614(\{s_1/s_0\}^2 T)^{1/5}$
Quadratic spectral	$b = 1.3221(\{s_1/s_0\}^2 T)^{1/5}$

Table 26.5 *continued*

Kernel Name	Bandwidth Parameter
Truncated	$b = 0.6611(\{s_1/s_0\}^2 T)^{1/5}$
Tukey-Hanning	$b = 1.7462(\{s_1/s_0\}^2 T)^{1/5}$

The terms s_0 and s_1 are estimated with the following formulas:

$$s_0 = \sigma_0 + 2\sum_{j=1}^{n}\sigma_j; \qquad s_1 = 2\sum_{j=1}^{n} j\sigma_j$$

where n is the lag selection parameter and is determined by kernels, as listed in Table 26.6.

Table 26.6 Lag Selection Parameter Estimation

Kernel Name	Lag Selection Parameter
Bartlett	$n = c(T/100)^{2/9}$
Parzen	$n = c(T/100)^{4/25}$
Quadratic spectral	$n = c(T/100)^{2/25}$
Truncated	$n = c(T/100)^{1/5}$
Tukey-Hanning	$n = c(T/100)^{1/5}$

The c in Table 26.6 is specified by the C= option; by default, C=12.

The σ_j is estimated with the equation

$$\sigma_j = T^{-1}\sum_{t=j+1}^{T}\left(\sum_{a=i}^{k} g_{at}\sum_{a=i}^{k} g_{at-j}\right), j = 0,\ldots,n$$

where g_{at} is the same as in the Andrews method and i is 1 if the NOINT option in the MODEL statement is specified, and 2 otherwise.

When you specify BANDWIDTH=SAMPLESIZE, the bandwidth parameter is estimated with the equation

$$b = \begin{cases} \lfloor \gamma T^r + c \rfloor & \text{if BANDWIDTH=SAMPLESIZE(INT) option is specified} \\ \gamma T^r + c & \text{otherwise} \end{cases}$$

where T is the sample size, $\lfloor x \rfloor$ is the largest integer less than or equal to x, and γ, r, and c are values specified by BANDWIDTH=SAMPLESIZE(GAMMA=, RATE=, CONSTANT=) options, respectively.

If the PREWHITENING option is specified in the MODEL statement, g_{it} is prewhitened by the VAR(1) model,

$$g_{it} = A_i g_{i,t-1} + w_{it}$$

Then Λ_{HAC} is calculated by

$$\Lambda_{\mathrm{HAC}} = a\sum_{i=1}^{N}\left\{\left(\sum_{t=1}^{T_i} w_{it}w_{it}' + \sum_{t=1}^{T_i}\sum_{s=1}^{t-1} k(\frac{s-t}{b})\left(w_{it}w_{is}' + w_{is}w_{it}'\right)\right)(I - A_i)^{-1}((I - A_i)^{-1})'\right\}$$

R-Square

The conventional R-square measure is inappropriate for all models that the PANEL procedure estimates by using GLS because a number outside the [0,1] range might be produced. Hence, a generalization of the R-square 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

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

However, the fixed-effects models are somewhat different. In the case of a fixed-effects model, the choice of including or excluding an intercept becomes merely a choice of classification. Suppressing the intercept in the FIXONE or FIXONETIME case merely changes the name of the intercept to a fixed effect. It makes no sense to redefine the R-square measure since nothing material changes in the model. Similarly, for the FIXTWO model there is no reason to change the R-square measure. In the case of the FIXONE, FIXONETIME, and FIXTWO models, the R-square is defined as the Theil (1961) R-square as shown in the preceding equation. This makes intuitive sense since you are regressing a transformed (demeaned) series on transformed regressors, excluding a constant. In other words, you are looking at 1 minus the sum of squared errors divided by the sum of squares of the (transformed) dependent variable.

In the case of OLS estimation, both of the R-square formulas given here reduce to the usual R-square formula.

Specification Tests

The PANEL procedure outputs the results of one specification test for fixed effects and two specification tests 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.

The Hausman (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

$$m = (\hat{\beta}_b - \hat{\beta}_a)'(\hat{\mathbf{S}}_b - \hat{\mathbf{S}}_a)^{-1}(\hat{\beta}_b - \hat{\beta}_a)$$

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. If the null hypothesis holds, the random-effects specification should be used.

Breusch and Pagan (1980) lay out a Lagrange multiplier test for random effects based on the simple OLS (pooled) estimator. If $\hat{u}_{it}$ is the itth residual from the OLS regression, then the Breusch-Pagan (BP) test for one-way random effects is

$$BP = \frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{N}\left[\sum_{t=1}^{T}\hat{u}_{it}\right]^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2} - 1\right]^2$$

The BP test generalizes to the case of a two-way random-effects model (Greene 2000, p. 589). Specifically,

$$\begin{aligned} BP2 &= \frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{n}\left[\sum_{t=1}^{T}\hat{u}_{it}\right]^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2} - 1\right]^2 \\ &+ \frac{NT}{2(N-1)}\left[\frac{\sum_{t=1}^{T}\left[\sum_{i=1}^{N}\hat{u}_{it}\right]^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2} - 1\right]^2 \end{aligned}$$

is distributed as a χ^2 statistic with two degrees of freedom. Since the BP2 test generalizes (nests the BP test) the test for random effects, the absence of random effects (nonrejection of the null of no random effects) in the BP2 is a fairly clear indication that there will probably not be any one-way effects either. In both cases (BP and BP2), the residuals are obtained from a pooled regression. There is very little extra cost in selecting both the BP and BP2 test. Notice that in the case of just groupwise heteroscedasticity, the BP2 test approaches BP. In the case of time based heteroscedasticity, the BP2 test reduces to a BP test of time effects. In the case of unbalanced panels, neither the BP nor BP2 statistics are valid.

Finally, you should be aware that the BP option generates different results depending on whether the estimation is FIXONE or FIXONETIME. Specifically, under the FIXONE estimation technique, the BP tests for cross-sectional random effects. Under the FIXONETIME estimation, the BP tests for time random effects.

While the Hausman statistic is automatically generated, you request Breusch-Pagan via the BP or BP2 option (see Baltagi 2008 for details).

Panel Data Poolability Test

The null hypothesis of poolability assumes homogeneous slope coefficients. An F test can be applied to test for the poolability across cross sections in panel data models.

F Test

For the unrestricted model, run a regression for each cross section and save the sum of squared residuals as SSE_u. For the restricted model, save the sum of squared residuals as SSE_r. If the test applies to all coefficients (including the constant), then the restricted model is the pooled model (OLS); if the test applies to coefficients other than the constant, then the restricted model is the fixed one-way model with cross-sectional fixed effects. If N and T denote the number of cross sections and time periods, then the number of observations is $n = NT$.[6] Let k be the number of regressors except the constant. The degree of freedom for the unrestricted model is $df_u = n - N(k+1)$. If the constant is restricted to be the same, the degree of freedom for the restricted model is $df_r = n - k - 1$ and the number of restrictions is $q = (N-1)(k+1)$. If the restricted model is the fixed one-way model, the degree of freedom is $df_r = n - k - N$ and the number of restrictions is $q = (N-1)k$. So the F test is

$$F = \frac{(SSE_r - SSE_u)/q}{SSE_u/df_u} \sim F(q, df_u)$$

For large N and T, you can use a chi-square distribution to approximate the limiting distribution, namely, $qF \implies \chi^2(q)$. The error term is assumed to be homogeneous; therefore, $\epsilon \sim \mathcal{N}\left(0, \sigma^2 I_n\right)$, and an OLS regression is sufficient. The test is the same as the Chow test (Chow 1960) extended to N linear regressions.

LR Test

Zellner (1962) also proved that the likelihood ratio test for null hypothesis of poolability can be based on the F statistic. The likelihood ratio can be expressed as $LR = -2log\left((1 + qF/df_u)^{-NT/2}\right) \implies LR = qF + O\left(n^{-1}\right)$. Under H_0, LR is asymptotically distributed as a chi-square with q degrees of freedom.

Panel Data Cross-Sectional Dependence Test

Breusch-Pagan LM Test

Breusch and Pagan (1980) propose a Lagrange multiplier (LM) statistic to test the null hypothesis of zero cross-sectional error correlations. Let e_{it} be the OLS estimate of the error term u_{it} under the null hypothesis. Then the pairwise cross-sectional correlations can be estimated by the sample counterparts $\hat{\rho}_{ij}$,

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t=\underline{T}_{ij}}^{\overline{T}_{ij}} e_{it} e_{jt}}{\sqrt{\sum_{t=\underline{T}_{ij}}^{\overline{T}_{ij}} e_{it}^2} \sqrt{\sum_{t=\underline{T}_{ij}}^{\overline{T}_{ij}} e_{jt}^2}}$$

where $\underline{T}_{ij}$ and $\overline{T}_{ij}$ are the lower bound and upper bound, respectively, which mark the overlap time periods for the cross sections i and j. If the panel is balanced, $\underline{T}_{ij} = 1$ and $\overline{T}_{ij} = T$. Let T_{ij} denote the number of

[6] For the unbalanced panel, the number of time series T_i might be different. The number of observations needs to be redefined accordingly.

overlapped time periods ($T_{ij} = \overline{T}_{ij} - \underline{T}_{ij} + 1$). Then the Breusch-Pagan LM test statistic can be constructed as

$$\mathrm{BP} = \sum_{i=1}^{N} \sum_{j=i+1}^{N} T_{ij} \hat{\rho}_{ij}^2$$

When N is fixed and $T_{ij} \to \infty$, $\mathrm{BP} \to \chi^2 \left(N \left(N-1 \right) /2 \right)$. So the test is not applicable as $N \to \infty$.

Because $\hat{\rho}_{ij}^2, i = 1, \ldots, N-1, j = i+1, \ldots, N$, are asymptotically independent under the null hypothesis of zero cross-sectional correlation, $T_{ij} \hat{\rho}_{ij}^2 \to \chi^2 (1)$. Then the following modified Breusch-Pagan LM statistic can be considered to test for cross-sectional dependence:

$$\mathrm{BP}s = \sqrt{\frac{1}{N \left(N-1 \right)}} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left(T_{ij} \hat{\rho}_{ij}^2 - 1 \right)$$

Under the null hypothesis, $\mathrm{BP}s \to \mathcal{N} (0, 1)$ as $T_{ij} \to \infty$, and then $N \to \infty$. But because $E \left(T_{ij} \hat{\rho}_{ij}^2 - 1 \right)$ is not correctly centered at zero for finite T_{ij}, the test is likely to exhibit substantial size distortion for large N and small T_{ij}.

Pesaran CD and CD*p* Test

Pesaran (2004) proposes a cross-sectional dependence test that is also based on the pairwise correlation coefficients $\hat{\rho}_{ij}$,

$$\mathrm{CD} = \sqrt{\frac{2}{N \left(N-1 \right)}} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \sqrt{T_{ij}} \hat{\rho}_{ij}$$

The test statistic has a zero mean for fixed N and T_{ij} under a wide class of panel data models, including stationary or unit root heterogeneous dynamic models that are subject to multiple breaks. For each $i \neq j$, as $T_{ij} \to \infty$, $\sqrt{T_{ij}} \hat{\rho}_{ij} \Longrightarrow \mathcal{N} (0, 1)$. Therefore, for N and T_{ij} tending to infinity in any order, $\mathrm{CD} \Longrightarrow \mathcal{N} (0, 1)$.

To enhance the power against the alternative hypothesis of local dependence, Pesaran (2004) proposes the CD*p* test. Local dependence is defined with respect to a weight matrix, $\mathbf{W} = \left(w_{ij} \right)$. Therefore, the test can be applied only if the cross-sectional units can be given an ordering that remains immutable over time. Under the alternative hypothesis of a *p*th-order local dependence, the CD statistic can be generalized to a local CD test, CD*p*,

$$\begin{aligned} \mathrm{CD}p &= \sqrt{\tfrac{2}{p(2N-p-1)}} \left(\textstyle\sum_{s=1}^{p} \sum_{i=s+1}^{N} \sqrt{T_{i,i-s}} \hat{\rho}_{i,i-s} \right) \\ &= \sqrt{\tfrac{2}{p(2N-p-1)}} \left(\textstyle\sum_{s=1}^{p} \sum_{i=1}^{N-s} \sqrt{T_{i,i+s}} \hat{\rho}_{i,i+s} \right) \end{aligned}$$

where $p = 1, \ldots, N-1$. When $p = N-1$, CD*p* reduces to the original CD test. Under the null hypothesis of zero cross-sectional dependence, the CD*p* statistic is centered at zero for fixed N and $T_{i,i-s} > k+1$, and $\mathrm{CD}p \Longrightarrow \mathcal{N} (0, 1)$ as $N \to \infty$ and $T_{i,i+s} \to \infty$.

Panel Data Unit Root Tests

Unit roots are a big concern in dynamic processes as they have important implications for the stationary of a process and hence estimation. Proceeding with regular estimation techniques ignoring the presence of units roots can lead to *spurious regressions* and hence produce nonsensical results. Therefore detecting unit roots to be able to analyze stationary processes is of vital concern for dynamic processes. One of the most widely used tests in the time series literature is the augmented Dickey-Fuller (ADF) test. This section introduces and briefly reviews the background information on the tests developed for dynamic panel data, which in most cases turn out to be enhancements of the ADF test.

Levin, Lin, and Chu (2002)

Levin, Lin, and Chu (2002) propose a panel unit root test for the null hypothesis of unit root against a homogeneous stationary hypothesis. The model is specified as

$$\Delta y_{it} = \delta y_{it-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{it-L} + \alpha_{mi} d_{mt} + \varepsilon_{it} \quad m = 1, 2, 3$$

The panel unit root test evaluates the null hypothesis of $H_0 : \delta = 0$, for all i, against the alternative hypothesis $H_1 : \delta < 0$ for all $i.$. Three models are considered: (1) $d_{1t} = \phi$ (the empty set) with no individual effects, (2) $d_{2t} = \{1\}$ in which the series y_{it} has an individual-specific mean but no time trend, and (3) $d_{3t} = \{1, t\}$ in which the series y_{it} has an individual-specific mean and linear and individual-specific time trend. The lag order p_i is unknown and is allowed to vary across individuals. It can be selected by the methods that are described in the section "Lag Order Selection in the ADF Regression" on page 1876. The selected lag order is denoted as $\hat{p}_i$. The necessary condition for the test is for $\frac{\sqrt{N}}{T} \rightarrow 0$. An important assumption is that the errors, ε_{it}, are assumed to be $i.i.d.(0, \sigma^2_{i,t})$. In other words, cross-sectional independence is assumed. The test is implemented in the following three steps:

Step 1 The ADF regressions are implemented for each individual i, and then the orthogonalized residuals are generated and normalized. That is, the following model is estimated:

$$\Delta y_{it} = \delta_i y_{it-1} + \sum_{L=1}^{\hat{p}_i} \theta_{iL} \Delta y_{it-L} + \alpha_{mi} d_{mt} + \varepsilon_{it} \quad m = 1, 2, 3$$

Then, two orthogonalized residuals are generated by the following two auxiliary regressions:

$$\Delta y_{it} = \sum_{L=1}^{\hat{p}_i} \theta_{iL} \Delta y_{it-L} + \alpha_{mi} d_{mi} + e_{it}$$

$$y_{it-1} = \sum_{L=1}^{\hat{p}_i} \theta_{iL} \Delta y_{it-L} + \alpha_{mi} d_{mi} + v_{it-1}$$

The residuals are then saved as $\hat{e}_{it}$ and $\hat{v}_{it-1}$, respectively, then normalized using the regression standard error from the ADF regression in order to remove heteroscedasticity. Let

$\hat{\sigma}_{\varepsilon i}$ denote the standard error from each of the previous ADF regressions, where $\hat{\sigma}^2_{\varepsilon i} = \sum_{t=\hat{p}_i+2}^{T} \left(\hat{e}_{it} - \hat{\delta}_i \hat{v}_{it-1}\right)^2 / (T - p_i - 1)$. The normalized residuals are then:

$$\tilde{e}_{it} = \frac{\hat{e}_{it}}{\hat{\sigma}_{\varepsilon i}}, \qquad \tilde{v}_{it-1} = \frac{\hat{v}_{it-1}}{\hat{\sigma}_{\varepsilon i}}$$

Step 2 The ratios of long-run to short-run standard deviations of Δy_{it} are estimated. Denote the ratios and the long-run variances as s_i and σ_{yi}, respectively. The long-run variances are estimated by the HAC (heteroscedasticity- and autocorrelation-consistent) estimators, which are described in the section "Long-Run Variance Estimation" on page 1876. Then the ratios are estimated by $\hat{s}_i = \hat{\sigma}_{yi}/\hat{\sigma}_{\varepsilon i}$. Let the average standard deviation ratio be $S_N = (1/N)\sum_{i=1}^{N} s_i$, and let its estimator be $\hat{S}_N = (1/N)\sum_{i=1}^{N} \hat{s}_i$. As the authors note in their paper, use of the long run variance based on first-differences results in lower bias in finite samples.

Step 3 The panel test statistics are calculated. To calculate the t statistic and the adjusted t statistic, the following equation is estimated:

$$\tilde{e}_{it} = \delta \tilde{v}_{it-1} + \tilde{\varepsilon}_{it}$$

The total number of observations is $N\tilde{T}$, with $\bar{\hat{p}} = \sum_{i=1}^{N} \hat{p}_i/N$, $\tilde{T} = T - \bar{\hat{p}} - 1$.
The standard t statistic for testing $H_0 : \delta = 0$ is $t_\delta = \hat{\delta}/\hat{\sigma}_\delta$, with OLS estimator $\hat{\delta}$ and standard deviation $\hat{\sigma}_\delta$.

$$\hat{\delta} = \frac{\sum_{i=1}^{N}\sum_{t=2+\hat{p}_i}^{T} \tilde{e}_{it}\tilde{v}_{it-1}}{\sum_{i=1}^{N}\sum_{t=2+\hat{p}_i}^{T} \tilde{v}^2_{it-1}}$$

$\hat{\sigma}_\delta = \hat{\sigma}_{\tilde{\varepsilon}}[\sum_{i=1}^{N}\sum_{t=2+\hat{p}_i}^{T} \tilde{v}^2_{it-1}]^{-\frac{1}{2}}$
Where $\hat{\sigma}_{\tilde{\varepsilon}}$ be the root mean square error from the step 3 regression

$$\hat{\sigma}^2_{\tilde{\varepsilon}} = \left[\frac{1}{N\tilde{T}} \sum_{i=1}^{N} \sum_{t=2+\hat{p}_i}^{T} (\tilde{e}_{it} - \hat{\delta}\tilde{v}_{it-1})^2\right]$$

However, the standard t statistic diverges to negative infinity for models (2) and (3). Levin, Lin, and Chu (2002) therefore propose the following adjusted t statistic:

$$t^*_\delta = \frac{t_\delta - N\tilde{T}\hat{S}_N\hat{\sigma}^{-2}_{\tilde{\varepsilon}}\hat{\sigma}_\delta\mu^*_{m\tilde{T}}}{\sigma^*_{m\tilde{T}}}$$

The mean and standard deviation adjustments ($\mu^*_{m\tilde{T}}, \sigma^*_{m\tilde{T}}$) depend on the time series dimension $\tilde{T}$ and model specification m, which can be found in Table 2 of Levin, Lin, and Chu (2002). The adjusted t statistic converges to the standard normal distribution. Therefore, the standard normal critical values are used in hypothesis testing.

Lag Order Selection in the ADF Regression

The methods for selecting the individual lag orders in the ADF regressions can be divided into two categories: selection based on information criteria and selection via sequential testing.

Lag Selection Based on Information Criteria In this method, the following information criteria can be applied to lag order selection: AIC, SBC, HQIC (HQC), and MAIC. As with other model selection applications, the lag order is selected from 0 to the maximum p_{max} to minimize the objective function, plus a penalty term, which is a function of the number of parameters in the regression. Let k be the number of parameters and T_o be the number of effective observations. For regression models, the objective function is $T_o log(SSR/T_o)$, where SSR is the sum of squared residuals. For AIC, the penalty term equals $2k$. For SBC, this term is $k log T_o$. For HQIC, it is $2ck log\,[log(T_o)]$ with c being a constant greater than 1.[7] For MAIC, the penalty term equals $2(\tau_T(k) + k)$, where

$$\tau_T(k) = (SSR/T_o)^{-1}\hat{\delta}^2 \sum_{t=p_{max}+2}^{T} y_{t-1}^2$$

and $\hat{\delta}$ is the estimated coefficient of the lagged dependent variable y_{t-1} in the ADF regression.

Lag Selection via Sequential Testing In this method, the lag order estimation is based on the statistical significance of the estimated AR coefficients. Hall (1994) proposed general-to-specific (GS) and specific-to-general (SG) strategies. Levin, Lin, and Chu (2002) recommend the first strategy, following Campbell and Perron (1991). In the GS modeling strategy, starting with the maximum lag order p_{max}, the t test for the largest lag order in $\hat{\theta}_i$ is performed to determine whether a smaller lag order is preferred. Specifically, when the null of $\hat{\theta}_{iL} = 0$ is not rejected given the significance level (5%), a smaller lag order is preferred. This procedure continues until a statistically significant lag order is reached. On the other hand, the SG modeling strategy starts with lag order 0 and moves toward the maximum lag order p_{max}.

Long-Run Variance Estimation

The long-run variance of Δy_{it} is estimated by a HAC-type estimator. For model (1), given the lag truncation parameter $\bar{K}$ and kernel weights $w_{\bar{K}L}$, the formula is

$$\hat{\sigma}_{yi}^2 = \frac{1}{T-1}\sum_{t=2}^{T}\Delta y_{it}^2 + 2\sum_{L=1}^{\bar{K}} w_{\bar{K}L}\left[\frac{1}{T-1}\sum_{t=2+L}^{T}\Delta y_{it}\Delta y_{it-L}\right]$$

To achieve consistency, the lag truncation parameter must satisfy $\bar{K}/T \to 0$ and $\bar{K} \to \infty$ as $T \to \infty$. Levin, Lin, and Chu (2002) suggest $\bar{K} = \left\lfloor 3.21T^{1/3} \right\rfloor$. The weights $w_{\bar{K}L}$ depend on the kernel function. Andrews (1991) proposes data-driven bandwidth (lag truncation parameter + 1 if integer-valued) selection procedures to minimize the asymptotic mean squared error (MSE) criterion. For more information about the kernel functions and Andrews (1991) data-driven bandwidth selection procedure, see the section "Heteroscedasticity- and Autocorrelation-Consistent Covariance Matrices" on page 1867. Because Levin, Lin, and Chu (2002) truncate the bandwidth as an integer, when LLCBAND is specified as the BANDWIDTH option, it corresponds to BANDWIDTH $= \left\lfloor 3.21T^{1/3} \right\rfloor + 1$. Furthermore, kernel weights $w_{\bar{K}L} = k(L/(\bar{K}+1))$ with kernel function $k(\cdot)$.

[7] In practice c is set to 1, following the literature (Hannan and Quinn 1979; Hall 1994).

For model (2), the series Δy_{it} is demeaned individual by individual first. Therefore, Δy_{it} is replaced by $\Delta y_{it} - \overline{\Delta y_{it}}$, where $\overline{\Delta y_{it}}$ is the mean of Δy_{it} for individual i. For model (3) with individual fixed effects and time trend, both the individual mean and trend should be removed before the long-run variance is estimated. That is, first regress Δy_{it} on $\{1, t\}$ for each individual and save the residual $\widetilde{\Delta y_{it}}$, and then replace Δy_{it} with the residual.

Cross-Sectional Dependence via Time-Specific Aggregate Effects

The Levin, Lin, and Chu (2002) testing procedure is based on the assumption of cross-sectional independence. It is possible to relax this assumption and allow for a limited degree of dependence via time-specific aggregate effects. Let θ_t denote the time-specific aggregate effects; then the data generating process (DGP) becomes

$$\Delta y_{it} = \delta y_{it-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{it-L} + \alpha_{mi} d_{mt} + \theta_t + \varepsilon_{it} \qquad m = 4, 5$$

Two more models are considered: (4) $d_{1t} = \phi$ (the empty set) with no individual effects, but with time effects, and (5) $d_{2t} = \{1\}$ in which the series y_{it} has an individual-specific mean and time-specific mean.

By subtracting the time averages $\bar{y}_t = \sum_{i=1}^{N} y_{it}$ from the observed dependent variable y_{it}, or equivalently, by including the time-specific intercepts θ_t in the ADF regression, the cross-sectional dependence is removed. The impact of a single aggregate common factor that has an identical impact on all individuals but changes over time can also be removed in this way. After cross-sectional dependence is removed, the three-step procedure is applied to calculate the Levin, Lin, and Chu (2002) adjusted t statistic.

Deterministic Variables

Three deterministic variables can be included in the model for the first-stage estimation: CS_FixedEffects (cross-sectional fixed effects), TS_FixedEffects (time series fixed effects), and TimeTrend (individual linear time trend). When a linear time trend is included, the individual fixed effects are also included. Otherwise the time trend is not identified. Moreover, if the time fixed effects are included, the time trend is not identified either. Therefore, we have 5 identified models: model (1), no deterministic variables; model (2), CS_FixedEffects; model (3), CS_FixedEffects and TimeTrend; model (4), TS_FixedEffects; model (5), CS_FixedEffects TS_FixedEffects. PROC PANEL outputs the test results for all 5 model specifications.

Im, Pesaran, and Shin (2003)

To test for the unit root in heterogeneous panels, Im, Pesaran, and Shin (2003) propose a standardized *t-bar* test statistic based on averaging the (augmented) Dickey-Fuller statistics across the groups. The limiting distribution is standard normal. The stochastic process y_{it} is generated by the first-order autoregressive process. If $\Delta y_{it} = y_{it} - y_{i,t-1}$, the data generating process can be expressed as in LLC:

$$\Delta y_{it} = \beta_i y_{it-1} + \sum_{j=1}^{p_i} \rho_{ij} \Delta y_{i,t-j} + \alpha_{mi} d_{mt} + \varepsilon_{it} \qquad m = 1, 2, 3$$

Unlike the DGP in LLC, β_i is allowed to differ across groups. The null hypothesis of unit roots is

$$H_0 : \beta_i = 0 \quad \text{for all } i$$

against the heterogeneous alternative,

$$H_1 : \beta_i < 0 \quad \text{for } i = 1, \ldots, N_1, \qquad \beta_i = 0 \quad \text{for } i = N_1 + 1, \ldots, N$$

The Im, Pesaran, and Shin test also allows for some (but not all) of the individual series to have unit roots under the alternative hypothesis. But the fraction of the individual processes that are stationary is positive, $\lim_{N\to\infty} N_1/N = \delta \in (0, 1]$. The *t-bar* statistic, denoted by $t\text{-}bar_{NT}$, is formed as a simple average of the individual t statistics for testing the null hypothesis of $\beta_i = 0$. If $t_{iT}\,(p_i, \rho_i)$ is the standard t statistic, then

$$t\text{-}bar_{NT} = N^{-1} \sum_{i=1}^{N} t_{iT}\,(p_i, \rho_i)$$

If $T \to \infty$, then for each i the t statistic (without time trend) converges to the Dickey-Fuller distribution, η_i, defined by

$$\eta_i = \frac{\frac{1}{2}\{[W_i(1)]^2 - 1\} - W_i(1) \int_0^1 W_i(u)du}{\int_0^1 [W_i(u)]^2 du - [\int_0^1 W_i(u)du]^2}$$

where W_i is the standard Brownian motion. The limiting distribution is different when a time trend is included in the regression (Hamilton 1994, p. 499). The mean and variance of the limiting distributions are reported in Nabeya (1999). The standardized *t-bar* statistic satisfies

$$Z_{t\text{-}bar}(p, \rho) = \frac{\sqrt{N}\{t\text{-}bar_{NT} - E(\eta)\}}{\sqrt{\mathrm{Var}(\eta)}} \implies \mathcal{N}(0, 1)$$

where the standard normal is the sequential limit with $T \to \infty$ followed by $N \to \infty$. To obtain better finite sample approximations, Im, Pesaran, and Shin (2003) propose standardizing the *t-bar* statistic by means and variances of $t_{iT}(p_i, 0)$ under the null hypothesis $\beta_i = 0$. The alternative standardized *t-bar* statistic is

$$W_{t\text{-}bar}(p, \rho) = \frac{\sqrt{N}\{t\text{-}bar_{NT} - N^{-1}\sum_{i=1}^{N} E[t_{iT}(p_i, 0) \mid \beta_i = 0]\}}{\{N^{-1}\sum_{i=1}^{N} \mathrm{Var}[t_{iT}(p_i, 0) \mid \beta_i = 0]\}^{1/2}} \implies \mathcal{N}(0, 1)$$

Im, Pesaran, and Shin (2003) simulate the values of $E[t_{iT}(p_i, 0) \mid \beta_i = 0]$ and $\mathrm{Var}[t_{iT}(p_i, 0) \mid \beta_i = 0]$ for different values of T and p. The lag order in the ADF regression can be selected by the same method as in Levin, Lin, and Chu (2002). For more information, see the section "Lag Order Selection in the ADF Regression" on page 1876.

When T is fixed, Im, Pesaran, and Shin (2003) assume serially uncorrelated errors, $p_i = 0$; t_{iT} is likely to have finite second moment, which is not established in the paper. The t statistic is modified by imposing the null hypothesis of a unit root. Denote $\tilde{\sigma}_{iT}$ as the estimated standard error from the restricted regression ($\beta_i = 0$),

$$\tilde{t}\text{-}bar_{NT} = N^{-1}\sum_{i=1}^{N} \tilde{t}_{it} = N^{-1}\sum_{i=1}^{N} \left[\hat{\beta}_{iT}\left(y'_{i,-1} M_\tau y_{i,-1}\right)^{1/2} / \tilde{\sigma}_{iT}\right]$$

where $\hat{\beta}_{iT}$ is the OLS estimator of β_i (unrestricted model), $\tau_T = (1, 1, \ldots, 1)'$, $M_\tau = I_T - \tau_T\left(\tau'_T \tau_T\right)\tau'_T$, and $y_{i,-1} = \left(y_{i0}, y_{i1}, \ldots, y_{i,T-1}\right)'$. Under the null hypothesis, the standardized $\tilde{t}\text{-}bar$ statistic converges to a standard normal variate,

$$Z_{\tilde{t}\text{-}bar} = \frac{\sqrt{N}\{\tilde{t}\text{-}bar_{NT} - E\left(\tilde{t}_T\right)\}}{\sqrt{\mathrm{Var}\left(\tilde{t}_T\right)}} \implies \mathcal{N}(0, 1)$$

where $E\left(\tilde{t}_T\right)$ and $\mathrm{Var}\left(\tilde{t}_T\right)$ are the mean and variance of $\tilde{t}_{iT}$, respectively. The limit is taken as $N \to \infty$ and T is fixed. Their values are simulated for finite samples without a time trend. The $Z_{\tilde{t}\text{-}bar}$ is also likely to converge to standard normal.

When N and T are both finite, an exact test that assumes no serial correlation can be used. The critical values of $t\text{-}bar_{NT}$ and $\tilde{t}\text{-}bar_{NT}$ are simulated.

Similar as in section "Levin, Lin, and Chu (2002)" on page 1874, it is possible to relax this assumption of cross-sectional independence and allow for a limited degree of dependence via time-specific aggregate effects. Two more models (model 4 and model 5) with time fixed effects are considered. For more information, see the section "Cross-Sectional Dependence via Time-Specific Aggregate Effects" on page 1877.

Combination Tests

Combining the observed significance levels (p-values) from N independent tests of the unit root null hypothesis was proposed by Maddala and Wu (1999); Choi (2001). Suppose G_i is the test statistic to test the unit root null hypothesis for individual $i = 1, \ldots, N$, and $F(\cdot)$ is the cdf (cumulative distribution function) of the asymptotic distribution as $T \to \infty$. Then the asymptotic p-value is defined as

$$p_i = F\left(G_i\right)$$

There are different ways to combine these p-values. The first one is the inverse chi-square test (Fisher 1932); this test is referred to as P test in Choi (2001) and λ in Maddala and Wu (1999):

$$P = -2\sum_{i=1}^{N} \ln\left(p_i\right)$$

When the test statistics $\{G_i\}_{i=1,\ldots,N}$ are continuous, $\{p_i\}_{i=1,\ldots,N}$ are independent uniform $(0, 1)$ variables. Therefore, $P \Rightarrow \chi^2_{2N}$ as $T \to \infty$ and N fixed. But as $N \to \infty$, P diverges to infinity in probability. Therefore, it is not applicable for large N. To derive a nondegenerate limiting distribution, the P test (Fisher test with $N \to \infty$) should be modified to

$$P_m = \sum_{i=1}^{N} \left(-2\ln\left(p_i\right) - 2\right) / 2\sqrt{N} = -\sum_{i=1}^{N} \left(\ln\left(p_i\right) + 1\right) / \sqrt{N}$$

Under the null as $T_i \to \infty$,[8] and then $N \to \infty$, $P_m \Rightarrow \mathcal{N}(0, 1)$.[9]

The second way of combining individual p-values is the inverse normal test,

$$Z = \sum_{i=1}^{N} \Phi^{-1}\left(p_i\right)$$

where $\Phi(\cdot)$ is the standard normal cdf. When $T_i \to \infty$, $Z \Rightarrow \mathcal{N}(0, 1)$ as N is fixed. When N and T_i are both large, the sequential limit is also standard normal if $T_i \to \infty$ first and $N \to \infty$ next.

The third way of combining p-values is the logit test,

$$L^* = \sqrt{k}L = \sqrt{k}\sum_{i=1}^{N} \ln\left(\frac{p_i}{1 - p_i}\right)$$

[8] The time series length T is subindexed by $i = 1, \ldots, N$ because the panel can be unbalanced.

[9] Choi (2001) also points out that the joint limit result where N and $\{T_i\}_{i=1,\ldots,N}$ go to infinity simultaneously is the same as the sequential limit, but it requires more moment conditions.

where $k = 3(5N+4)/\left(\pi^2 N(5N+2)\right)$. When $T_i \to \infty$ and N is fixed, $L^* \Rightarrow t_{5N+4}$. In other words, the limiting distribution is the t distribution with degree of freedom $5N+4$. The sequential limit is $L^* \Rightarrow \mathcal{N}(0,1)$ as $T_i \to \infty$ and then $N \to \infty$. Simulation results in Choi (2001) suggest that the Z test outperforms other combination tests. For the time series unit root test G_i, Maddala and Wu (1999) apply the augmented Dickey-Fuller test. According to Choi (2006), the Elliott, Rothenberg, and Stock (1996) Dickey-Fuller generalized least squares (DF-GLS) test brings significant size and power advantages in finite samples.

Similar as in section "Levin, Lin, and Chu (2002)" on page 1874, it is possible to relax this assumption of cross-sectional independence and allow for a limited degree of dependence via time-specific aggregate effects. Two more models (model 4 and model 5) with time fixed effects are considered. For more information, see the section "Cross-Sectional Dependence via Time-Specific Aggregate Effects" on page 1877.

Breitung's Unbiased Tests

To account for the nonzero mean of the t statistic in the OLS detrending case, bias-adjusted t statistics were proposed by: Levin, Lin, and Chu (2002); Im, Pesaran, and Shin (2003). The bias corrections imply a severe loss of power. Breitung and associates take an alternative approach to avoid the bias, by using alternative estimates of the deterministic terms (Breitung and Meyer 1994; Breitung 2000; Breitung and Das 2005). The DGP is the same as in the Im, Pesaran, and Shin approach. When serial correlation is absent, for model (2) with individual specific means, the constant terms are estimated by the initial values y_{i1}. Therefore, the series y_{it} is adjusted by subtracting the initial value. The equation becomes

$$\Delta y_{it} = \delta^* \left(y_{i,t-1} - y_{i1}\right) + v_{it}$$

For model (3) with individual specific means and time trends, the time trend can be estimated by $\hat{\beta}_i = (T-1)^{-1}(y_{iT} - y_{i1})$. The levels can be transformed as

$$\tilde{y}_{it} = y_{it} - y_{i1} - \hat{\beta}_i t = y_{it} - y_{i1} - t(y_{iT} - y_{i1})/(T-1)$$

The Helmert transformation is applied to the dependent variable to remove the mean of the differenced variable:

$$\Delta y^*_{it} = \sqrt{\frac{T-t}{T-t+1}}\left[\Delta y_{it} - \left(\Delta y_{i,t+1} + \cdots + \Delta y_{iT}\right)/(T-t)\right]$$

The transformed model is

$$\Delta y^*_{it} = \delta^* \tilde{y}_{i,t-1} + v_{it}$$

The pooled t statistic has a standard normal distribution. Therefore, no adjustment is needed for the t statistic. To adjust for heteroscedasticity across cross sections, Breitung (2000) proposes a UB (unbiased) statistic based on the transformed data,

$$UB = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T} \Delta y^*_{it}\tilde{y}_{i,t-1}/\sigma_i^2}{\sqrt{\sum_{i=1}^{N}\sum_{t=2}^{T} \tilde{y}^2_{i,t-1}/\sigma_i^2}}$$

where $\sigma_i^2 = E(\Delta y_{it} - \beta_i)^2$. When σ_i^2 is unknown, it can be estimated as

$$\hat{\sigma}_i^2 = \sum_{t=2}^{T}\left(\Delta y_{it} - \sum_{t=2}^{T}\Delta y_{it}/(T-1)\right)^2/(T-2)$$

The UB statistic has a standard normal limiting distribution as $T \to \infty$ followed by $N \to \infty$ sequentially. To account for the short-run dynamics, Breitung and Das (2005) suggest applying the test to the prewhitened series, $\hat{y}_{it}$. For model (1) and model (2) (constant-only case), they suggested the same method as in step 1 of Levin, Lin, and Chu (2002).[10] For model (3) (with a constant and linear time trend), the prewhitened series can be obtained by running the following restricted ADF regression under the null hypothesis of a unit root ($\delta = 0$) and no intercept and linear time trend ($\mu_i = 0, \beta_i = 0$):

$$\Delta y_{it} = \sum_{L=1}^{\hat{p}_i} \theta_{iL} \Delta y_{it-L} + \mu_i + \varepsilon_{it}$$

where $\hat{p}_i$ is a consistent estimator of the true lag order p_i and can be estimated by the procedures listed in the section "Lag Order Selection in the ADF Regression" on page 1876. For LLC and IPS tests, the lag orders are selected by running the ADF regressions. But for Breitung and his coauthors' tests, the restricted ADF regressions are used to be consistent with the prewhitening method. Let $\left(\hat{\mu}_i, \hat{\theta}_{iL}\right)$ be the estimated coefficients.[11] The prewhitened series can be obtained by

$$\Delta \hat{y}_{it} = \Delta y_{it} - \sum_{L=1}^{\hat{p}_i} \hat{\theta}_{iL} \Delta y_{it-L}$$

and

$$\hat{y}_{it} = y_{it} - \sum_{L=1}^{\hat{p}_i} \hat{\theta}_{iL} y_{it-L}$$

The transformed series are random walks under the null hypothesis,

$$\Delta \hat{y}_{it} = \delta \hat{y}_{i,t-1} + v_{it}$$

where $y_{is} = 0$ for $s < 0$. When the cross-section units are independent, the t statistic converges to standard normal under the null, as $T \to \infty$ followed by $N \to \infty$,

$$t_{OLS} = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} y_{i,t-1} \Delta y_{it}}{\hat{\sigma} \sqrt{\sum_{i=1}^{N} \sum_{t=2}^{T} y_{i,t-1}^2}} \implies \mathcal{N}(0,1)$$

where $\hat{\sigma}^2 = \sum_{i=1}^{N} \sum_{t=2}^{T} \left(\Delta y_{it} - \hat{\delta} y_{i,t-1}\right)^2 / N\,(T-1)$ with OLS estimator $\hat{\delta}$.
To take account for cross-sectional dependence, Breitung and Das (2005) propose the robust t statistic and a GLS version of the test statistic. Let $v_t = (v_{1t}, \ldots, v_{Nt})'$ be the error vector for time t, and let $\Omega = E\left(v_t v_t'\right)$ be a positive definite matrix with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_N$. Let $y_t = (y_{1t}, \ldots, y_{Nt})'$ and $\Delta y_t = (\Delta y_{1t}, \ldots, \Delta y_{Nt})'$. The model can be written as a SUR-type system of equations,

$$\Delta y_t = \delta y_{t-1} + v_t$$

[10] For more information, see the section "Levin, Lin, and Chu (2002)" on page 1874. The only difference is the standard error estimate $\hat{\sigma}_{\varepsilon i}^2$. Breitung suggests using $T - p_i - 2$ instead of $T - p_i - 1$ as in LLC to normalize the standard error.

[11] Breitung (2000) suggests the approach in step 1 of Levin, Lin, and Chu (2002), while Breitung and Das (2005) suggest the prewhitening method as described above. In Breitung's code, to be consistent with the papers, different approaches are adopted for model (2) and (3). Meanwhile, for the order of variable transformation and prewhitening, in model (2), the initial values are deducted (variable transformation) first, and then the prewhitening was applied. For model (3), the order is reversed. The series is prewhitened and then transformed to remove the mean and linear time trend.

The unknown covariance matrix Ω can be estimated by its sample counterpart,

$$\hat{\Omega} = \sum_{t=2}^{T} \left(\Delta y_t - \hat{\delta} y_{t-1}\right) \left(\Delta y_t - \hat{\delta} y_{t-1}\right)' / (T-1)$$

The sequential limit $T \to \infty$ followed by $N \to \infty$ of the standard t statistic t_{OLS} is normal with mean 0 and variance $v_\Omega = \lim_{N\to\infty} \mathrm{tr}\left(\Omega^2/N\right) / (\mathrm{tr}\Omega/N)^2$. The variance v_Ω can be consistently estimated by $\hat{v}_{\hat{\delta}} = \left(\sum_{t=2}^{T} y'_{t-1} \hat{\Omega} y_{t-1}\right) / \left(\sum_{t=2}^{T} y'_{t-1} y_{t-1}\right)^2$. Thus the robust t statistic can be calculated as

$$t_{rob} = \frac{\delta}{\hat{v}_{\hat{\delta}}} = \frac{\sum_{t=2}^{T} y'_{t-1} \Delta y_t}{\sqrt{\sum_{t=2}^{T} y'_{t-1} \hat{\Omega} y_{t-1}}} \implies \mathcal{N}(0,1)$$

as $T \to \infty$ followed by $N \to \infty$ under the null hypothesis of random walk. Since the finite sample distribution can be quite different, Breitung and Das (2005) list the 1%, 5%, and 10% critical values for different N's.

When $T > N$, a (feasible) GLS estimator is applied; it is asymptotically more efficient than the OLS estimator. The data are transformed by multiplying $\hat{\Omega}^{-1/2}$ as defined before, $\hat{z}_t = \hat{\Omega}^{-1/2} y_t$. Thus the model is transformed into

$$\Delta \hat{z}_t = \delta \hat{z}_{t-1} + e_t$$

The feasible GLS (FGLS) estimator of δ and the corresponding t statistic are obtained by estimating the transformed model by OLS and denoted by $\hat{\delta}_{GLS}$ and t_{GLS}, respectively:

$$t_{GLS} = \frac{\sum_{t=2}^{T} y'_{t-1} \hat{\Omega}^{-1} \Delta y_t}{\sqrt{\sum_{t=2}^{T} y'_{t-1} \hat{\Omega}^{-1} y_{t-1}}} \implies \mathcal{N}(0,1)$$

Similar as in section "Levin, Lin, and Chu (2002)" on page 1874, it is possible to relax this assumption of cross-sectional independence and allow for a limited degree of dependence via time-specific aggregate effects. Two more models (model 4 and model 5) with time fixed effects are considered. For more information, see the section "Cross-Sectional Dependence via Time-Specific Aggregate Effects" on page 1877.

Hadri (2000) Stationarity Tests

Hadri (2000) adopts a component representation where an individual time series is written as a sum of a deterministic trend, a random walk, and a white-noise disturbance term. Under the null hypothesis of stationary, the variance of the random walk equals 0. Specifically, two models are considered:

- For model (1), the time series y_{it} is stationary around a level r_{i0},

$$y_{it} = r_{it} + \epsilon_{it} \quad i = 1, \ldots, N, \quad t = 1, \ldots, T$$

- For model (2), y_{it} is trend stationary,

$$y_{it} = r_{it} + \beta_i t + \epsilon_{it} \quad i = 1, \ldots, N, \quad t = 1, \ldots, T$$

where r_{it} is the random walk component,

$$r_{it} = r_{it-1} + u_{it} \quad i = 1, \ldots, N, \quad t = 1, \ldots, T$$

The initial values of the random walks, $\{r_{i0}\}_{i=1,\ldots,N}$, are assumed to be fixed unknowns and can be considered as heterogeneous intercepts. The errors ϵ_{it} and u_{it} satisfy $\epsilon_{it} \sim \text{iid}\mathcal{N}\left(0, \sigma_\epsilon^2\right)$, $u_{it} \sim \text{iid}\mathcal{N}\left(0, \sigma_u^2\right)$ and are mutually independent.

The null hypothesis of stationarity is $H_0 : \sigma_u^2 = 0$ against the alternative random walk hypothesis $H_1 : \sigma_u^2 > 0$.

In matrix form, the models can be written as

$$y_i = X_i \beta_i + e_i$$

where $y_i' = (y_{i1}, \ldots, y_{iT})$, $e_i' = (e_{i1}, \ldots, e_{iT})$ with $e_{it} = \sum_{j=1}^{t} u_{ij} + \epsilon_{it}$, and $X_i = (\iota_T, a_T)$ with ι_T being a $T \times 1$ vector of ones, $a_T' = (1, \ldots, T)$, and $\beta_i' = (r_{i0}, \beta_i)$.

Let $\hat{\epsilon}_{it}$ be the residuals from the regression of y_i on X_i; then the LM statistic is

$$LM = \frac{\frac{1}{N}\sum_{i=1}^{N} \frac{1}{T^2} \sum_{t=1}^{T} S_{it}^2}{\hat{\sigma}_\epsilon^2}$$

where $S_{it} = \sum_{j=1}^{t} \hat{\epsilon}_{ij}$ is the partial sum of the residuals and $\hat{\sigma}_\epsilon^2$ is a consistent estimator of σ_ϵ^2 under the null hypothesis of stationarity. With some regularity conditions,

$$LM \xrightarrow{p} E\left[\int_0^1 V^2(r)\, dr\right]$$

where $V(r)$ is a standard Brownian bridge in model (1) and a second-level Brownian bridge in model (2). Let $W(r)$ be a standard Wiener process (Brownian motion),

$$V(r) = \begin{cases} W(r) - rW(1) & \text{for model (1)} \\ W(r) + \left(2r - 3r^2\right) W(1) + 6r(r-1) \int_0^1 W(s)\, ds & \text{for model (2)} \end{cases}$$

The mean and variance of the random variable $\int V^2$ can be calculated by using the characteristic functions,

$$\xi = E\left[\int_0^1 V^2(r)\, dr\right] = \begin{cases} \frac{1}{6} & \text{for model (1)} \\ \frac{1}{15} & \text{for model (2)} \end{cases}$$

and

$$\zeta^2 = var\left[\int_0^1 V^2(r)\, dr\right] = \begin{cases} \frac{1}{45} & \text{for model (1)} \\ \frac{11}{6300} & \text{for model (2)} \end{cases}$$

The LM statistics can be standardized to obtain the standard normal limiting distribution,

$$Z = \frac{\sqrt{N}(LM - \xi)}{\zeta} \Longrightarrow \mathcal{N}(0, 1)$$

Consistent Estimator of σ_ϵ^2

Hadri's (2000) test can be applied to the general case of heteroscedasticity and serially correlated disturbance errors. Under homoscedasticity and serially uncorrelated errors, σ_ϵ^2 can be estimated as

$$\hat{\sigma}_\epsilon^2 = \sum_{i=1}^{N}\sum_{t=1}^{T} \hat{\epsilon}_{it}^2 / N\,(T-k)$$

where k is the number of regressors. Therefore, $k = 1$ for model (1) and $k = 2$ for model (2).

When errors are heteroscedastic across individuals, the standard errors $\sigma_{\epsilon,i}^2$ can be estimated by $\hat{\sigma}_{\epsilon,i}^2 = \sum_{t=1}^{T} \hat{\epsilon}_{it}^2 / (T-k)$ for each individual i and the LM statistic needs to be modified to

$$LM = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\frac{1}{T^2}\sum_{t=1}^{T} S_{it}^2}{\hat{\sigma}_{\epsilon,i}^2}\right)$$

To allow for temporal dependence over t, σ_ϵ^2 has to be replaced by the long-run variance of ϵ_{it}, which is defined as $\sigma^2 = \sum_{i=1}^{N} \lim_{T\to\infty} T^{-1}\left(S_{iT}^2\right)/N$. A HAC estimator can be used to consistently estimate the long-run variance σ^2. For more information, see the section "Long-Run Variance Estimation" on page 1876.

Similar as in section "Levin, Lin, and Chu (2002)" on page 1874, it is possible to relax this assumption of cross-sectional independence and allow for a limited degree of dependence via time-specific aggregate effects. One more models (model 3) with time fixed effects are considered. For more information, see the section "Cross-Sectional Dependence via Time-Specific Aggregate Effects" on page 1877.

Harris and Tzavalis (1999) Panel Unit Root Tests

Harris and Tzavalis (1999) derive the panel unit root test under fixed T and large N. Five models are considered as in Levin, Lin, and Chu (2002). Model (1) is the homogeneous panel,

$$y_{it} = \varphi y_{it-1} + v_{it}$$

Under the null hypothesis, $\varphi = 1$. For model (2), each series is a unit root process with a heterogeneous drift,

$$y_{it} = \alpha_i + \varphi y_{it-1} + v_{it}$$

Model (3) includes heterogeneous drifts and linear time trends,

$$y_{it} = \alpha_i + \beta_i t + \varphi y_{it-1} + v_{it}$$

Similar as in section "Levin, Lin, and Chu (2002)" on page 1874, it is possible to relax this assumption of cross-sectional independence and allow for a limited degree of dependence via time-specific aggregate effects. Two more models (model 4 and model 5) with time fixed effects are considered. For more information, see the section "Cross-Sectional Dependence via Time-Specific Aggregate Effects" on page 1877.

Let $\hat{\varphi}$ be the OLS estimator of φ; then

$$\hat{\varphi} - 1 = \left[\sum_{i=1}^{N} y_{i,-1}' Q_T y_{i,-1}\right]^{-1} \cdot \left[\sum_{i=1}^{N} y_{i,-1}' Q_T v_i\right]$$

where $y_{i,-1} = (y_{i0}, \ldots, y_{iT-1})$, $v_i' = (v_{i1}, \ldots, v_{iT})$, and Q_T is the projection matrix. For model (1), there are no regressors other than the lagged dependent value, so Q_T is the identity matrix I_T. For model

(2), a constant is included, so $Q_T = I_T - e_T e_T'/T$ with e_T a $T \times 1$ column of ones. For model (3), a constant and time trend are included. Thus $Q_T = I_T - Z_T \left(Z_T' Z_T\right)^{-1} Z_T'$, where $Z_T = (e_T, \tau_T)$ and $\tau_T = (1, \ldots, T)'$.

When $y_{i0} = 0$ in model (1) under the null hypothesis, as $N \to \infty$

$$\sqrt{NT\,(T-1)/2}\,(\hat{\varphi} - 1) \xrightarrow{y_{i0}=0, H_0} \mathcal{N}(0,1)$$

As $T \to \infty$, it becomes $T\sqrt{N}\,(\hat{\varphi} - 1) \xrightarrow{H_0} \mathcal{N}(0,2)$.

When the drift is absent in model (2), $\alpha_i = 0$, under the null hypothesis, as $N \to \infty$

$$\sqrt{\frac{5N\,(T+1)^3\,(T-1)}{3\left(17T^2 - 20T + 17\right)}}\left(\hat{\varphi} - 1 + \frac{3}{(T+1)}\right) \xrightarrow{\alpha_i=0, H_0} \mathcal{N}(0,1)$$

As $T \to \infty$, $\left(T\sqrt{N}\,(\hat{\varphi} - 1) + 3\sqrt{N}\right)/\sqrt{51/5} \xrightarrow{H_0} \mathcal{N}(0,1)$.

When the time trend is absent in model (3), $\beta_i = 0$, under the null hypothesis, as $N \to \infty$

$$\sqrt{\frac{112N\,(T+2)^3\,(T-2)}{15\left(193T^2 - 728T + 1147\right)}})\left(\hat{\varphi} - 1 + \frac{15}{2\,(T+2)}\right) \xrightarrow{\beta_i=0, H_0} \mathcal{N}(0,1)$$

When $T \to \infty$, $\left(T\sqrt{N}\,(\hat{\varphi} - 1) + 7.5\sqrt{N}\right)/\sqrt{2895/112} \xrightarrow{H_0} \mathcal{N}(0,1)$.

Lagrange Multiplier (LM) Tests for Cross-Sectional and Time Effects

For random one-way and two-way error component models, the Lagrange multiplier test for the existence of cross-sectional or time effects or both is based on the residuals from the restricted model (that is, the pooled model). For more information about the Breusch-Pagan LM test, see the section "Specification Tests" on page 1870.

Honda (1985) and Honda (1991) UMP Test and Moulton and Randolph (1989) SLM Test

The Breusch-Pagan LM test is two-sided when the variance components are nonnegative. For a one-sided alternative hypothesis, Honda (1985) suggests a uniformly most powerful (UMP) LM test for $H_0^1 : \sigma_\gamma^2 = 0$ (no cross-sectional effects) that is based on the pooled estimator. The alternative is the one-sided $H_1^1 : \sigma_\gamma^2 > 0$. Let $\hat{u}_{it}$ be the residual from the simple pooled OLS regression and $d = \left(\sum_{i=1}^N \left[\sum_{t=1}^T \hat{u}_{it}\right]^2\right)/\left(\sum_{i=1}^N \sum_{t=1}^T \hat{u}_{it}^2\right)$. Then the test statistic is defined as

$$J \equiv \sqrt{\frac{NT}{2(T-1)}}\,[d-1] \xrightarrow{H_0^1} \mathcal{N}(0,1)$$

The square of J is equivalent to the Breusch and Pagan (1980) LM test statistic. Moulton and Randolph (1989) suggest an alternative standardized Lagrange multiplier (SLM) test to improve the asymptotic approximation for Honda's one-sided LM statistic. The SLM test's asymptotic critical values are usually closer to the exact

critical values than are those of the LM test. The SLM test statistic standardizes Honda's statistic by its mean and standard deviation. The SLM test statistic is

$$\mathrm{S} \equiv \frac{J - E(J)}{\sqrt{\mathrm{Var}(J)}} = \frac{d - E(d)}{\sqrt{\mathrm{Var}(d)}} \rightarrow \mathcal{N}(0,1)$$

Let $D = I_N \bigotimes J_T$, where J_T is the $T \times T$ square matrix of 1s. The mean and variance can be calculated by the formulas

$$E(d) = \mathrm{Tr}(DM_Z)/(n-k)$$

$$\mathrm{Var}(d) = 2\{(n-k)\mathrm{Tr}\,(DM_Z)^2 - [\mathrm{Tr}\,(DM_Z)]^2\}/((n-k)^2(n-k+2))$$

where Tr denotes the trace of a particular matrix, Z represents the regressors in the pooled model, $n = NT$ is the number of observations, k is the number of regressors, and $M_Z = I_n - Z(Z'Z)^{-1}Z'$. To calculate $\mathrm{Tr}(DM_Z)$, let $Z = \left(Z_1', Z_2', \ldots, Z_N\right)'$. Then

$$\mathrm{Tr}(DM_Z) = NT - \mathrm{Tr}\left(J_T \sum_{i=1}^{N}\left[Z_i\left(\sum_{j=1}^{N} Z_j' Z_j\right)^{-1} Z_i'\right]\right)$$

To test for $H_0^2 : \sigma_\alpha^2 = 0$ (no time effects), define $d2 = \left(\sum_{t=1}^{T}\left[\sum_{i=1}^{N}\hat{u}_{it}\right]^2\right) / \left(\sum_{t=1}^{T}\sum_{i=1}^{N}\hat{u}_{it}^2\right)$. Then the test statistic is modified as

$$J2 \equiv \sqrt{\frac{NT}{2(N-1)}}\,[d2 - 1] \xrightarrow{H_0^2} \mathcal{N}(0,1)$$

$J2$ can be standardized by $D = J_N \bigotimes I_T$, and other parameters are unchanged. Therefore,

$$\mathrm{S2} \equiv \frac{J2 - E(J2)}{\sqrt{\mathrm{Var}(J2)}} = \frac{d2 - E(d2)}{\sqrt{\mathrm{Var}(d2)}} \rightarrow \mathcal{N}(0,1)$$

To test for $H_0^3 : \sigma_\gamma^2 = 0, \sigma_\alpha^2 = 0$ (no cross-sectional and time effects), the test statistic is $J3 = (J + J2)/\sqrt{2}$ and $D = \sqrt{n/(T-1)}\,(I_N \bigotimes J_T)/2 + \sqrt{n/(N-1)}\,(J_N \bigotimes I_T)/2$. To standardize, define $d3 = \sqrt{n/(T-1)}d/2 + \sqrt{n/(N-1)}(d2)/2$,

$$\mathrm{S3} \equiv \frac{J3 - E(J3)}{\sqrt{\mathrm{Var}(J3)}} = \frac{d3 - E(d3)}{\sqrt{\mathrm{Var}(d3)}} \rightarrow \mathcal{N}(0,1)$$

King and Wu (1997) LMMP Test and the SLM Test

King and Wu (1997) derive the locally mean most powerful (LMMP) one-sided test for H_0^1 and H_0^2, which coincides with the Honda (1985) UMP test. Baltagi, Chang, and Li (1992) extend the King and Wu (1997) test for H_0^3 as follows:

$$\mathrm{KW} \equiv \frac{\sqrt{T-1}}{\sqrt{N+T-2}}\mathrm{J} + \frac{\sqrt{N-1}}{\sqrt{N+T-2}}\mathrm{J2} \xrightarrow{H_0^3} \mathcal{N}(0,1)$$

For the standardization, use $D = I_N \bigotimes J_T + J_N \bigotimes I_T$. Define $d_{kw} = d + d2$; then

$$\mathrm{S_{kw}} \equiv \frac{KW - E(KW)}{\sqrt{\mathrm{Var}(KW)}} = \frac{d_{kw} - E(d_{kw})}{\sqrt{\mathrm{Var}(d_{kw})}} \rightarrow \mathcal{N}(0,1)$$

Gourieroux, Holly, and Monfort (1982) LM Test

If one or both variance components (σ_γ^2 and σ_α^2) are small and close to 0, the test statistics J and $J2$ can be negative. Baltagi, Chang, and Li (1992) follow Gourieroux, Holly, and Monfort (1982) and propose a one-sided LM test for H_0^3, which is immune to the possible negative values of J and $J2$. The test statistic is

$$\text{GHM} \equiv \begin{cases} J^2 + (J2)^2 & \text{if } J > 0,\ J2 > 0 \\ J^2 & \text{if } J > 0,\ J2 \leq 0 \\ (J2)^2 & \text{if } J \leq 0,\ J2 > 0 \\ 0 & \text{if } J \leq 0,\ J2 \leq 0 \end{cases} \xrightarrow{H_0^3} \left(\frac{1}{4}\right)\chi^2(0) + \left(\frac{1}{2}\right)\chi^2(1) + \left(\frac{1}{4}\right)\chi^2(2)$$

where $\chi^2(0)$ is the unit mass at the origin.

Tests for Serial Correlation and Cross-Sectional Effects

The presence of cross-sectional effects causes serial correlation in the errors. Therefore, serial correlation is often tested jointly with cross-sectional effects. Joint and conditional tests for both serial correlation and cross-sectional effects have been covered extensively in the literature.

Baltagi and Li Joint LM Test for Serial Correlation and Random Cross-Sectional Effects

Baltagi and Li (1991) derive the LM test statistic, which jointly tests for zero first-order serial correlation and random cross-sectional effects under normality and homoscedasticity. The test statistic is independent of the form of serial correlation, so it can be used with either AR(1) or MA(1) error terms. The null hypothesis is a white noise component: $H_0^1 : \sigma_\gamma^2 = 0, \theta = 0$ for MA(1) with MA coefficient θ or $H_0^2 : \sigma_\gamma^2 = 0, \rho = 0$ for AR(1) with AR coefficient ρ. The alternative is either a one-way random-effects model (cross-sectional) or first-order serial correlation AR(1) or MA(1) in errors or both. Under the null hypothesis, the model can be estimated by the pooled estimation (OLS). Denote the residuals as $\hat{u}_{it}$. The test statistic is

$$\text{BL91} = \frac{NT^2}{2(T-1)(T-2)}\left[A^2 - 4AB + 2TB^2\right] \xrightarrow{H_0^{1,2}} \chi^2(2)$$

where

$$A = \frac{\sum_{i=1}^N \left(\sum_{t=1}^T \hat{u}_{it}\right)^2}{\sum_{i=1}^N \sum_{t=1}^T \hat{u}_{it}^2} - 1, \qquad B = \frac{\sum_{i=1}^N \sum_{t=2}^T \hat{u}_{it}\hat{u}_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T \hat{u}_{it}^2}$$

Wooldridge Test for the Presence of Unobserved Effects

Wooldridge (2002, sec. 10.4.4) suggests a test for the absence of an unobserved effect. Under the null hypothesis $H_0 : \sigma_\gamma^2 = 0$, the errors u_{it} are serially uncorrelated. To test $H_0 : \sigma_\gamma^2 = 0$, Wooldridge (2002) proposes to test for AR(1) serial correlation. The test statistic that he proposes is

$$W = \frac{\sum_{i=1}^N \sum_{t=1}^{T-1} \sum_{s=t+1}^T \hat{u}_{it}\hat{u}_{is}}{\left[\sum_{i=1}^N \left(\sum_{t=1}^{T-1} \sum_{s=t+1}^T \hat{u}_{it}\hat{u}_{is}\right)^2\right]^{1/2}} \to \mathcal{N}(0,1)$$

where $\hat{u}_{it}$ are the pooled OLS residuals. The test statistic W can detect many types of serial correlation in the error term u, so it has power against both the one-way random-effects specification and the serial correlation in error terms.

Bera, Sosa Escudero, and Yoon Modified Rao's Score Test in the Presence of Local Misspecification

Bera, Sosa Escudero, and Yoon (2001) point out that the standard specification tests, such as the Honda (1985) test described in the section "Honda (1985) and Honda (1991) UMP Test and Moulton and Randolph (1989) SLM Test" on page 1885, are not valid when they test for either cross-sectional random effects or serial correlation without considering the presence of the other effects. They suggest a modified Rao's score (RS) test. When A and B are defined as in Baltagi and Li (1991), the test statistic for testing serial correlation under random cross-sectional effects is

$$\mathrm{RS}^{*}_{\rho} = \frac{NT^2\,(B - A/T)^2}{(T-1)\,(1-2/T)}$$

Baltagi and Li (1991, 1995) derive the conventional RS test when the cross-sectional random effects is assumed to be absent:

$$\mathrm{RS}_{\rho} = \frac{NT^2B^2}{T-1}$$

Symmetrically, to test for the cross-sectional random effects in the presence of serial correlation, the modified Rao's score test statistic is

$$\mathrm{RS}^{*}_{\mu} = \frac{NT\,(A-2B)^2}{2\,(T-1)\,(1-2/T)}$$

and the conventional Rao's score test statistic is given in Breusch and Pagan (1980). The test statistics are asymptotically distributed as $\chi^2\,(1)$.

Because $\sigma^2_{\gamma} > 0$, the one-sided test is expected to lead to more powerful tests. The one-sided test can be derived by taking the signed square root of the two-sided statistics:

$$\mathrm{RSO}^{*}_{\mu} = \sqrt{\frac{NT}{2\,(T-1)\,(1-2/T)}}\,(A-2B) \rightarrow \mathcal{N}\,(0,1)$$

Baltagi and Li (1995) LM Test for First-Order Correlation under Fixed Effects

The two-sided LM test statistic for testing a white noise component in a fixed one-way model ($H^5_0 : \theta = 0$ or $H^6_0 : \rho = 0$, given that γ_i are fixed effects) is

$$\mathrm{BL95} = \frac{NT^2}{T-1}\left(\frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{u}_{it}\hat{u}_{i,t-1}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}^2_{it}}\right)^2$$

where $\hat{u}_{it}$ are the residuals from the fixed one-way model (FIXONE). The LM test statistic is asymptotically distributed as χ^2_1 under the null hypothesis. The one-sided LM test with alternative hypothesis $\rho > 0$ is

$$\mathrm{BL95}_2 = \sqrt{\frac{NT^2}{T-1}}\,\frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\hat{u}_{it}\hat{u}_{i,t-1}}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}^2_{it}}$$

which is asymptotically distributed as standard normal.

Durbin-Watson Statistic

Bhargava, Franzini, and Narendranathan (1982) propose a test of the null hypothesis of no serial correlation ($H_0^6 : \rho = 0$) against the alternative ($H_1^6 : 0 < |\rho| < 1$) by the Durbin-Watson statistic,

$$d_\rho = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\left(\hat{u}_{it} - \hat{u}_{i,t-1}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2}$$

where $\hat{u}_{it}$ are the residuals from the fixed one-way model (FIXONE).

The test statistic d_ρ lies somewhere between 0 and 4, inclusive where $d_\rho = 2$ indicates no serial correlation. Values closer to 0 indicate positive serial correlation while values closer to 4 indicate negative serial correlation. To test against a positive correlation ($\rho > 0$) for very large N, you can simply detect whether $d_\rho < 2$. However, for small to moderate N, the mechanics of the Durbin-Watson test produce an indeterminate region, a region of uncertainty as to whether to reject the null hypothesis. The output contains two p-values: The first, Pr < DWLower, treats the uncertainty region as a rejection region. The second, Pr > DWUpper, is more conservative and treats the uncertainty region as a failure-to-reject region. You can think of these two p-values as bounds on the exact p-value. Some of the upper and lower bounds are listed in Bhargava, Franzini, and Narendranathan (1982). The test statistic d_ρ is a locally most powerful invariant test in the neighborhood of $\rho = 0$.

Berenblut-Webb Statistic

Bhargava, Franzini, and Narendranathan (1982) also suggest using the Berenblut-Webb statistic, which is a locally most powerful invariant test in the neighborhood of $\rho = 1$. The test statistic is

$$\mathrm{g}_\rho = \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta\tilde{u}_{i,t}^2}{\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{u}_{it}^2}$$

where $\Delta\tilde{u}_{it}$ are the residuals from the first-difference estimation. The upper and lower bounds are the same as for the Durbin-Watson statistic d_ρ and produce two p-values, one conservative and one anti-conservative.

Testing for Random Walk Null Hypothesis

You can also use the Durbin-Watson and Berenblut-Webb statistics to test the random walk null hypothesis, with the bounds that are listed in Bhargava, Franzini, and Narendranathan (1982). For more information about these statistics, see the sections "Durbin-Watson Statistic" on page 1889 and "Berenblut-Webb Statistic" on page 1889. Bhargava, Franzini, and Narendranathan (1982) also propose the R_ρ statistic to test the random walk null hypothesis $\rho = 1$ against the stationary alternative $|\rho| < 1$. Let $\mathbf{F}^* = I_N \otimes \mathbf{F}$, where $\mathbf{F}$ is a $(T-1)(T-1)$ symmetric matrix that has the following elements:

$$\mathbf{F}_{tt'} = \left(T - t'\right)t/T \quad if\, t' \geq t \quad \left(t, t' = 1, \ldots, T-1\right)$$

The test statistic is

$$\begin{aligned}\mathrm{R}_\rho &= \Delta\tilde{U}'\Delta\tilde{U}/\Delta\tilde{U}'F^*\Delta\tilde{U} \\ &= \frac{\sum_{i=1}^{N}\sum_{t=2}^{T}\Delta\tilde{u}_{i,t}^2}{\left[\sum_{i=1}^{N}\sum_{t=2}^{T}(t-1)(T-t+1)\Delta\tilde{u}_{i,t}^2 + 2\sum_{i=1}^{N}\sum_{t=2}^{T-1}\sum_{t'=t+1}^{T}(T-t'+1)(t-1)\Delta\tilde{u}_{i,t}\Delta\tilde{u}_{i,t'}\right]/T}\end{aligned}$$

The statistics R_ρ, g_ρ, and d_ρ can be used with the same bounds. They satisfy $\mathrm{R}_\rho \leq \mathrm{g}_\rho \leq \mathrm{d}_\rho$, and they are equivalent for large panels.

Troubleshooting

You need to follow some guidelines when you use PROC PANEL for analysis. For each cross section, PROC PANEL requires at least two time series observations that have nonmissing values for all model variables. There should be at least two cross sections for each time point in the data. If these two conditions are not met, then an error message is printed in the log that states that there is only one cross section or time series observation and further computations will be terminated. You must provide adequate data for an estimation method to produce results, and you should check the log for any errors that are related to data.

If PROC PANEL uses the Parks method and the number of cross sections is greater than the number of time series observations per cross section, then PROC PANEL produces an error message that states that the ϕ matrix is singular. This is analogous to seemingly unrelated regression that has fewer observations than equations in the model. To avoid the problem, reduce the number of cross sections.

Your data set could have multiple observations for each time ID within a particular cross section. However, you can use PROC PANEL only in cases where you have only a single observation for each time ID within each cross section. In such a case, after you have sorted the data, an error warning is printed in the log that states that the data have not been sorted in ascending sequence with respect to time series ID.

The cause of the error is due to multiple observations for each time ID for a given cross section. PROC PANEL allows only one observation for each time ID within each cross section.

The data set shown in Figure 26.2 illustrates the preceding instance with the correct representation.

Figure 26.2 Single Observation for Each Time Series

Obs	firm	year	production	cost
1	1	1955	5.36598	1.14867
2	1	1960	6.03787	1.45185
3	1	1965	6.37673	1.52257
4	1	1970	6.93245	1.76627
5	2	1955	6.54535	1.35041
6	2	1960	6.69827	1.71109
7	2	1965	7.40245	2.09519
8	2	1970	7.82644	2.39480

In this case, you can observe that there are no multiple observations with respect to a given time series ID within a cross section. This is the correct representation of a data set where PROC PANEL is applicable.

If for state ID 1 you have two observations for the year=1955, then PROC PANEL produces the following error message:

"The data set is not sorted in ascending sequence with respect to time series ID. The current time period has year=1955 and the previous time period has year=1955 in cross section firm=1."

A data set similar to the previous example with multiple observations for the YEAR=1955 is shown in Figure 26.3; this data set results in an error message due to multiple observations while using PROC PANEL.

Figure 26.3 Multiple Observations for Each Time Series

Obs	firm	year	production	cost
1	1	1955	5.36598	1.14867
2	1	1955	6.37673	1.52257
3	1	1960	6.03787	1.45185
4	1	1970	6.93245	1.76627
5	2	1955	6.54535	1.35041
6	2	1960	6.69827	1.71109
7	2	1965	7.40245	2.09519
8	2	1970	7.82644	2.39480

In order to use PROC PANEL, you need to aggregate the data so that you have unique time ID values within each cross section. One possible way to do this is to run PROC MEANS on the input data set and compute the mean of all the variables by FIRM and YEAR, and then use the output data set.

Creating ODS Graphics

Statistical procedures use ODS Graphics to create graphs as part of their output. ODS Graphics is described in detail in Chapter 21, "Statistical Graphics Using ODS" (*SAS/STAT User's Guide*).

Before you create graphs, ODS Graphics must be enabled (for example, with the ODS GRAPHICS ON statement). For more information about enabling and disabling ODS Graphics, see the section "Enabling and Disabling ODS Graphics" in that chapter.

The overall appearance of graphs is controlled by ODS styles. Styles and other aspects of using ODS Graphics are discussed in the section "A Primer on ODS Statistical Graphics" in that chapter.

This section describes the use of ODS for creating graphics with the PANEL procedure. The table below lists the graph names, the plot descriptions, and the options used.

Table 26.7 ODS Graphics Produced by PROC PANEL

ODS Graph Name	Plot Description	PLOTS= Options
DiagnosticsPanel	All applicable plots listed below	
ResidualPlot	Plot of the residuals	RESIDUAL, RESID
FitPlot	Predicted versus actual plot	FITPLOT
QQPlot	Plot of the quantiles of the residuals	QQ
ResidSurfacePlot	Surface plot of the residuals	RESIDSURFACE
PredSurfacePlot	Surface plot of the predicted values	PREDSURFACE
ActSurfacePlot	Surface plot of the actual values	ACTSURFACE
ResidStackPlot	Stack plot of the residuals	RESIDSTACK, RESSTACK
ResidHistogram	Plot of the histogram of residuals	RESIDUALHISTOGRAM, RESIDHISTOGRAM

OUTPUT OUT= Data Set

PROC PANEL writes the initial data of the estimated model, predicted values, and residuals to an output data set when the OUTPUT OUT= statement is specified. The OUT= data set contains the following variables:

MODELL is a character variable that contains the label for the MODEL statement if a label is specified.

METHOD is a character variable that identifies the estimation method.

MODLNO is the number of the model estimated.

ACTUAL contains the value of the dependent variable.

WEIGHT contains the weighing variable.

CSID is the value of the cross section ID.

TSID is the value of the time period in the dynamic model.

regressors are the values of regressor variables specified in the MODEL statement.

name if PRED=*name1* and/or RESIDUAL=*name2* options are specified, then *name1* and *name2* are the columns of predicted values of dependent variable and residuals of the regression, respectively.

OUTEST= Data Set

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

MODEL is a character variable that contains the label for the MODEL statement if a label is specified.

METHOD is a character variable that identifies the estimation method.

TYPE is a character variable that identifies the type of observation. Values of the _TYPE_ variable are CORRB, COVB, CSPARMS, STD, and the type of model estimated. 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, the STD observation indicates the row of standard deviations of the corresponding coefficients, and the type of model estimated observation contains the parameter estimates.

NAME is a character variable that contains the name of a regressor variable for COVB and CORRB observations and is 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 is a character variable that contains the name of the response variable.

MSE is the mean square error of the transformed model.

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

VARCS is the variance component estimate due to cross sections. The _VARCS_ variable is included in the OUTEST= data set when a one-way or two-way random-effects models is estimated.

VARTS is the variance component estimate due to time series. The _VARTS_ variable is included in the OUTEST= data set when a two-way random-effects model is estimated.

VARERR is the variance component estimate due to error. The _VARERR_ variable is included in the OUTEST= data set when a one-way or two-way random-effects models is estimated.

_A_1 is the first-order autoregressive parameter estimate. The _A_1 variable 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. The _A_1 variable has a value only for _TYPE_=CSPARMS observations. The cross section to which the estimate belongs is indicated by the _CSID_ variable.

Intercept is the intercept parameter estimate. (Intercept is missing for models when the NOINT option is specified.)

regressors are 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.

OUTTRANS= Data Set

PROC PANEL writes the transformed series to an output data set. That is, if the user selects FIXONE, FIXONETIME, FDONE, FDONETIME, or RANONE and supplies the OUTTRANS= option, the transformed dependent variable and independent variables are written out to a SAS data set; other variables in the input data set are copied unchanged.

Suppose your data set contains the variables y, x1, x2, x3, and z2. The following statements result in a SAS data set:

```
proc panel data=datain outtrans=dataout;
   id cs ts;
   model y = x1 x2 x3 / fixone;
run;
```

First, z2 is copied over. Then _Int, x1, x2, y, and x3 are replaced with their mean deviates (from cross sections). Furthermore, two new variables are created.

MODELL is the model's label (if it exists).

METHOD is the model's transformation type. _METHOD_ reflects the estimation method and, in the case of random effects, the variance-component method.

Printed Output

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

- 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. For fixed-effects model analysis, an F test for the absence of fixed effects is produced, and for random-effects model analysis, a Hausman test is used for the appropriateness of the random-effects specification.
- the estimates of the underlying error structure parameters
- 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.

Optionally, PROC PANEL prints the following:

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

ODS Table Names

PROC PANEL 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 Table 26.8.

Table 26.8 ODS Tables Produced in PROC PANEL

ODS Table Name	Description	Options
ODS Tables Created by the MODEL Statement		
ModelDescription	Model description	Default
FitStatistics	Fit statistics	Default
FixedEffectsTest	F test for no fixed effects	FIXONE, FIXTWO, FIXONETIME
ParameterEstimates	Parameter estimates	Default
CovB	Covariances of parameter estimates	COVB

Table 26.8 *continued*

ODS Table Name	Description	Options
CorrB	Correlations of parameter estimates	CORRB
VarianceComponents	Variance component estimates	RANONE, RANTWO, DASILVA
RandomEffectsTest	Hausman test for random effects	RANONE, RANTWO
HausmanTest	Hausman specification test	HTAYLOR, AMACURDY
AR1Estimates	First-order autoregressive parameter estimates	RHO(PARKS)
BFNTest	R_ρ statistic for serial correlation	BFN
BL91Test	Baltagi and Li joint LM test	BL91
BL95Test	Baltagi and Li (1995) LM test	BL95
BreuschPaganTest	Breusch-Pagan one-way test	BP
BreuschPaganTest2	Breusch-Pagan two-way test	BP2
BSYTest	Bera, Sosa Escudero, and Yoon modified RS test	BSY
BWTest	Berenblut-Webb statistic for serial correlation	BW
DWTest	Durbin-Watson statistic for serial correlation	DW
GHMTest	Gourieroux, Holly, and Monfort two-way test	GHM
HondaTest	Honda one-way test	HONDA
HondaTest2	Honda two-way test	HONDA2
KingWuTest	King and Wu two-way test	KW
WOOLDTest	Wooldridge (2002) test for unobserved effects	WOOLDRIDGE02
CDTestResults	Cross-sectional dependence test	CDTEST
CDpTestResults	Local cross-sectional dependence test	CDTEST
Sargan	Sargan's test for overidentification	GMM1, GMM2, ITGMM
ARTest	Autoregression test for the residuals	GMM1, GMM2, ITGMM
IterHist	Iteration history	ITPRINT(ITGMM)
ConvergenceStatus	Convergence status of iterated GMM estimator	ITGMM
EstimatedPhiMatrix	Estimated phi matrix	PARKS
EstimatedAutocovariances	Estimates of autocovariances	DASILVA
LLCResults	LLC panel unit root test	UROOTTEST
IPSResults	IPS panel unit root test	UROOTTEST
CTResults	Combination test for panel unit root	UROOTTEST
HadriResults	Hadri panel stationarity test	UROOTTEST
HTResults	Harris and Tzavalis panel unit root test	UROOTTEST
BRResults	Breitung panel unit root test	UROOTTEST
URootdetail	Panel unit root test intermediate results	UROOTTEST
PTestResults	Poolability test for panel data	POOLTEST

Table 26.8 *continued*

ODS Table Name	Description	Options
ODS Tables Created by the COMPARE Statement		
StatComparisonTable	Comparison of model fit statistics	
ParameterComparisonTable	Comparison of model parameter estimates, standard errors, and *t* tests	
ODS Tables Created by the TEST Statement		
TestResults	Test results	

Examples: PANEL Procedure

Example 26.1: 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. See Feige (1964) for data description. All variables were transformed via natural logarithm. The data set A is shown below.

```
data a;
   length state $ 2;
   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;
CA   1949  6.2785  6.1924  4.4998  7.2056 -1.0700  0.1080  1.0664
CA   1950  6.4019  6.2106  4.6821  7.2889 -1.0106  0.1501  1.0767
CA   1951  6.5058  6.2729  4.8598  7.3827 -1.0024  0.4008  1.1291
CA   1952  6.4785  6.2729  5.0039  7.4000 -0.9970  0.4492  1.1227
CA   1953  6.4118  6.2538  5.1761  7.4200 -0.8916  0.4662  1.2110
CA   1954  6.4520  6.2971  5.3613  7.4478 -0.6951  0.4756  1.1924

   ... more lines ...
```

As shown in the following statements, the SORT procedure is used to sort the data into the required time series cross-sectional format; then PROC PANEL analyzes the data:

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

proc panel data=a;
   model d = y rd rt rs / rantwo vcomp = fb 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 26.1.1, Output 26.1.2, and Output 26.1.3, 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. 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 (see Output 26.1.2, Output 26.1.3, and Output 26.1.5). Time deposits are also substitutes for S&L shares in the time deposit demand equation (see Output 26.1.4), while these liquid assets are independent of each other in Output 26.1.5 (insignificant coefficient estimate of RT, −0.02705). Demand deposits and time deposits appear to be weak complements in Output 26.1.3 and Output 26.1.4, while the cross elasticities between demand deposits and time deposits are not significant in Output 26.1.2 and Output 26.1.5.

Output 26.1.1 Demand for Demand Deposits, Fuller-Battese Variance Component with Two-Way Random–Effects Model

The PANEL Procedure
Fuller and Battese Variance Components (RanTwo)

Dependent Variable: d (Per Capita Demand Deposits)

Model Description	
Estimation Method	RanTwo
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			
Coefficients	DF	m Value	Pr > m
4	4	5.51	0.2385

Output 26.1.1 *continued*

Parameter Estimates						
Variable	DF	Estimate	Standard 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 26.1.2 Demand for Demand Deposits, Parks Method

The PANEL 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	40.0198	DFE	72
MSE	0.5558	Root MSE	0.7455
R-Square	0.9263		

Parameter Estimates						
Variable	DF	Estimate	Standard Error	t Value	Pr > |t|	Label
Intercept	1	-2.66565	0.4250	-6.27	<.0001	Intercept
y	1	1.222569	0.0573	21.33	<.0001	Permanent Per Capita Personal Income
rd	1	-0.43591	0.0272	-16.03	<.0001	Service Charge on Demand Deposits
rt	1	0.041237	0.0284	1.45	0.1505	Interest on Time Deposits
rs	1	-0.26683	0.0886	-3.01	0.0036	Interest on S & L Association Shares

Output 26.1.3 Demand for Demand Deposits, DaSilva Method

The PANEL 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

Output 26.1.3 *continued*

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

Parameter Estimates

Variable	DF	Estimate	Standard 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 26.1.4 Demand for Time Deposits, Parks Method

The PANEL 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	34.5713	**DFE**	72
MSE	0.4802	**Root MSE**	0.6929
R-Square	0.9517		

Output 26.1.4 *continued*

Parameter Estimates						
Variable	DF	Estimate	Standard Error	t Value	Pr > |t|	Label
Intercept	1	-5.33334	0.6780	-7.87	<.0001	Intercept
y	1	1.516344	0.1097	13.82	<.0001	Permanent Per Capita Personal Income
rd	1	-0.04791	0.0399	-1.20	0.2335	Service Charge on Demand Deposits
rt	1	0.227152	0.0449	5.06	<.0001	Interest on Time Deposits
rs	1	-0.42569	0.1708	-2.49	0.0150	Interest on S & L Association Shares

Output 26.1.5 Demand for Savings and Loan Shares, Parks Method

The PANEL 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	39.2550	DFE	72
MSE	0.5452	Root MSE	0.7384
R-Square	0.9017		

Parameter Estimates						
Variable	DF	Estimate	Standard Error	t Value	Pr > |t|	Label
Intercept	1	-8.09632	1.0628	-7.62	<.0001	Intercept
y	1	1.832988	0.1567	11.70	<.0001	Permanent Per Capita Personal Income
rd	1	0.576723	0.0589	9.80	<.0001	Service Charge on Demand Deposits
rt	1	-0.02705	0.0423	-0.64	0.5242	Interest on Time Deposits
rs	1	0.408066	0.1478	2.76	0.0073	Interest on S & L Association Shares

Example 26.2: The Airline Cost Data: Fixtwo Model

The Christenson Associates airline data are a frequently cited data set (see Greene 2000). The data measure costs, prices of inputs, and utilization rates for six airlines over the time span 1970–1984. This example analyzes the log transformations of the cost, price and quantity, and the raw (not logged) capacity utilization measure. You speculate the following model,

$$\begin{aligned} ln\,(TC_{it}) = \alpha_N + \gamma_T \quad &+ \quad (\alpha_i - \alpha_N) + (\gamma_t - \gamma_T) \\ &+ \quad \beta_1 ln\,(Q_{it}) + \beta_2 ln\,(PF_{it}) + \beta_3 LF_{it} + \epsilon_{it} \end{aligned}$$

where the α are the pure cross-sectional effects and γ are the time effects. The actual model speculated is highly nonlinear in the original variables. It would look like the following:

$$TC_{it} = exp\left(\alpha_i + \gamma_t + \beta_3 LF_{it} + \epsilon_{it}\right) Q_{it}^{\beta_1} PF_{it}^{\beta_2}$$

The data and preliminary SAS statements are as follows:

```
data airline;
   input  Obs I T C Q PF LF;
   label obs = "Observation number";
   label I  = "Firm Number (CSID)";
   label T  = "Time period (TSID)";
   label Q  = "Output in revenue passenger miles (index)";
   label C  = "Total cost, in thousands";
   label PF = "Fuel price";
   label LF = "Load Factor (utilization index)";
datalines;
 1    1     1    1140640    0.95276     106650    0.53449
 2    1     2    1215690    0.98676     110307    0.53233
 3    1     3    1309570    1.09198     110574    0.54774
 4    1     4    1511530    1.17578     121974    0.54085

   ... more lines ...


data airline;
   set airline;
   lC = log(C);
   lQ = log(Q);
   lPF = log(PF);
   label lC = "Log Transformation of Costs";
   label lQ = "Log Transformation of Quantity";
   label lPF= "Log Transformation of Price of Fuel";
run;
```

The following statements fit the model:

```
proc panel data=airline printfixed;
   id i t;
   model lC = lQ lPF LF / fixtwo;
run;
```

First, you see the model's description in Output 26.2.1. The model is a two-way fixed-effects model. There are six cross sections and fifteen time observations.

Output 26.2.1 The Airline Cost Data—Model Description

The PANEL Procedure
Fixed Two-Way Estimates

Dependent Variable: lC (Log Transformation of Costs)

Model Description	
Estimation Method	FixTwo
Number of Cross Sections	6
Time Series Length	15

The R-square and degrees of freedom can be seen in Table 26.2.2. On the whole, you see a large R-square, so there is a reasonable fit. The degrees of freedom of the estimate are 90 minus 14 time dummy variables minus 5 cross section dummy variables and 4 regressors.

Output 26.2.2 The Airline Cost Data—Fit Statistics

Fit Statistics			
SSE	0.1768	**DFE**	67
MSE	0.0026	**Root MSE**	0.0514
R-Square	0.9984		

The F test for fixed effects is shown in Table 26.2.3. Testing the hypothesis that there are no fixed effects, you easily reject the null of poolability. There are group effects, or time effects, or both. The test is highly significant. OLS would not give reasonable results.

Output 26.2.3 The Airline Cost Data—Test for Fixed Effects

F Test for No Fixed Effects			
Num DF	**Den DF**	**F Value**	**Pr > F**
19	67	23.10	<.0001

Looking at the parameters, you see a more complicated pattern. Most of the cross-sectional effects are highly significant (with the exception of CS2). This means that the cross sections are significantly different from the sixth cross section. Many of the time effects show significance, but this is not uniform. It looks like the significance might be driven by a large 16^{th} period effect, since the first six time effects are negative and of similar magnitude. The time dummy variables taper off in size and lose significance from time period 12 onward. There are many causes to which you could attribute this decay of time effects. The time period of the data spans the OPEC oil embargoes and the dissolution of the Civil Aeronautics Board (CAB). These two forces are two possible reasons to observe the decay and parameter instability. As for the regression parameters, you see that quantity affects cost positively, and the price of fuel has a positive effect, but load factors negatively affect the costs of the airlines in this sample. The somewhat disturbing result is that the fuel cost is not significant. If the time effects are proxies for the effect of the oil embargoes, then an insignificant fuel cost parameter would make some sense. If the dummy variables proxy for the dissolution of the CAB, then the effect of load factors is also not being precisely estimated.

Output 26.2.4 The Airline Cost Data—Parameter Estimates

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Pr > |t|	Label
CS1	1	0.174237	0.0861	2.02	0.0470	Cross Sectional Effect 1
CS2	1	0.111412	0.0780	1.43	0.1576	Cross Sectional Effect 2
CS3	1	-0.14354	0.0519	-2.77	0.0073	Cross Sectional Effect 3
CS4	1	0.18019	0.0321	5.61	<.0001	Cross Sectional Effect 4
CS5	1	-0.04671	0.0225	-2.08	0.0415	Cross Sectional Effect 5
TS1	1	-0.69286	0.3378	-2.05	0.0442	Time Series Effect 1
TS2	1	-0.63816	0.3321	-1.92	0.0589	Time Series Effect 2
TS3	1	-0.59554	0.3294	-1.81	0.0751	Time Series Effect 3
TS4	1	-0.54192	0.3189	-1.70	0.0939	Time Series Effect 4
TS5	1	-0.47288	0.2319	-2.04	0.0454	Time Series Effect 5
TS6	1	-0.42705	0.1884	-2.27	0.0267	Time Series Effect 6
TS7	1	-0.39586	0.1733	-2.28	0.0255	Time Series Effect 7
TS8	1	-0.33972	0.1501	-2.26	0.0269	Time Series Effect 8
TS9	1	-0.2718	0.1348	-2.02	0.0478	Time Series Effect 9
TS10	1	-0.22734	0.0763	-2.98	0.0040	Time Series Effect 10
TS11	1	-0.1118	0.0319	-3.50	0.0008	Time Series Effect 11
TS12	1	-0.03366	0.0429	-0.78	0.4354	Time Series Effect 12
TS13	1	-0.01775	0.0363	-0.49	0.6261	Time Series Effect 13
TS14	1	-0.01865	0.0305	-0.61	0.5430	Time Series Effect 14
Intercept	1	12.93834	2.2181	5.83	<.0001	Intercept
lQ	1	0.817264	0.0318	25.66	<.0001	Log Transformation of Quantity
lPF	1	0.168732	0.1635	1.03	0.3057	Log Transformation of Price of Fuel
LF	1	-0.88267	0.2617	-3.37	0.0012	Load Factor (utilization index)

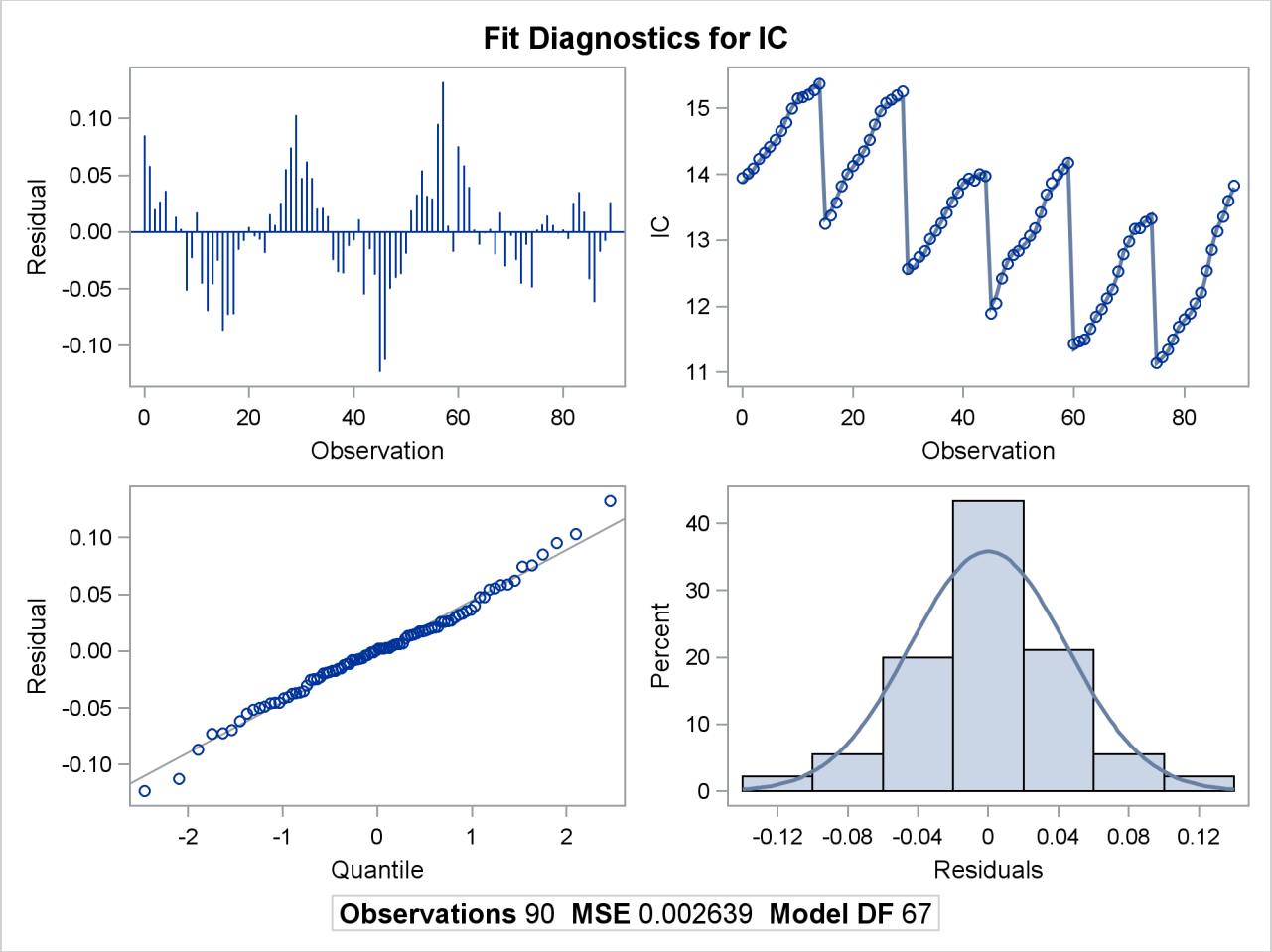
ODS Graphics Plots

ODS graphics plots can be obtained to graphically analyze the results. The following statements show how to generate the plots. If the PLOTS=ALL option is specified, all available plots are produced in two panels. For a complete list of options, see the section "Creating ODS Graphics" on page 1891.

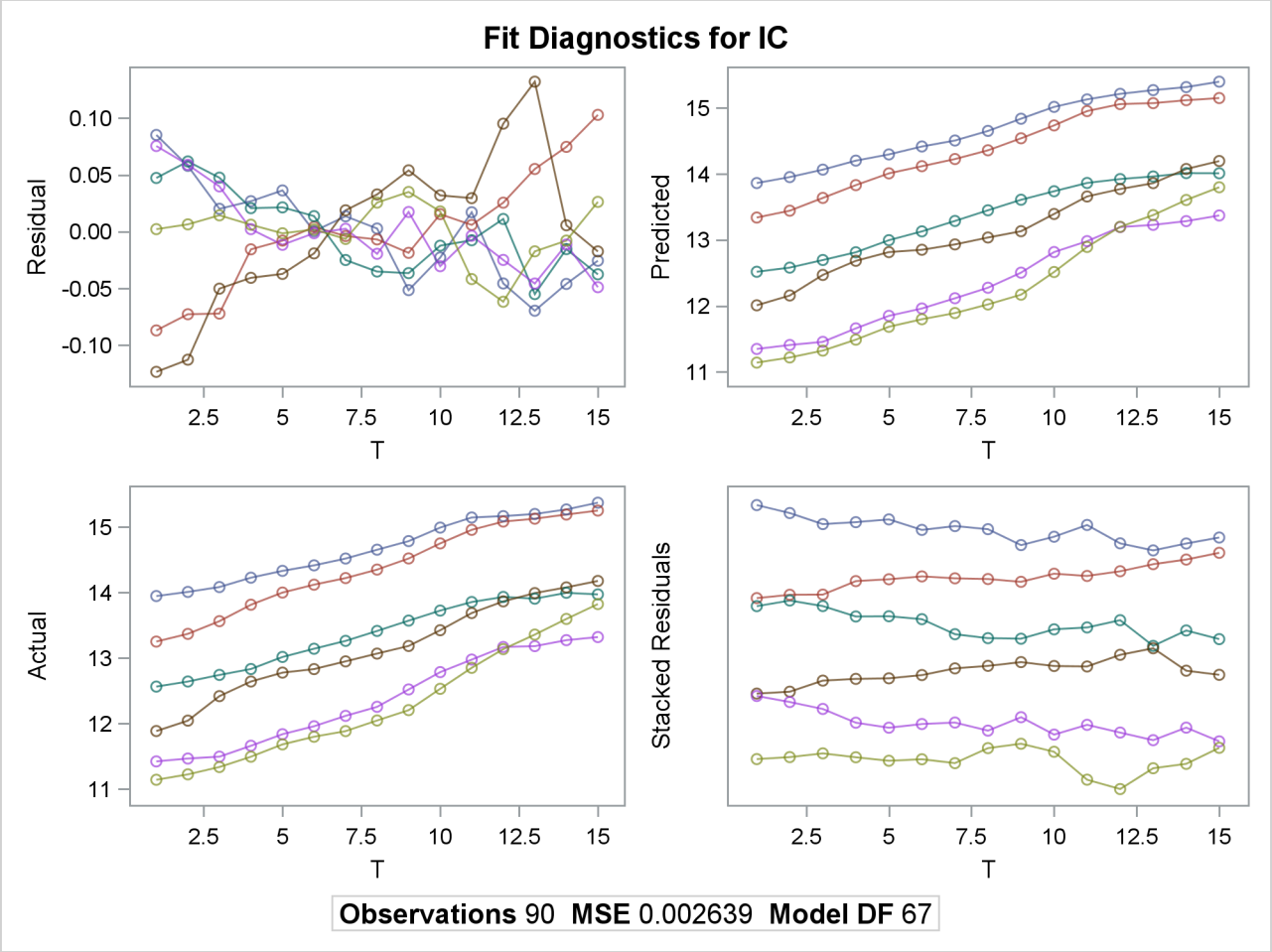
```
ods graphics on;
proc panel data=airline;
   id i t;
   model lC = lQ lPF LF / fixtwo plots = all;
run;
```

The preceding statements result in plots shown in Output 26.2.5 and Output 26.2.6.

Output 26.2.5 Diagnostic Panel 1



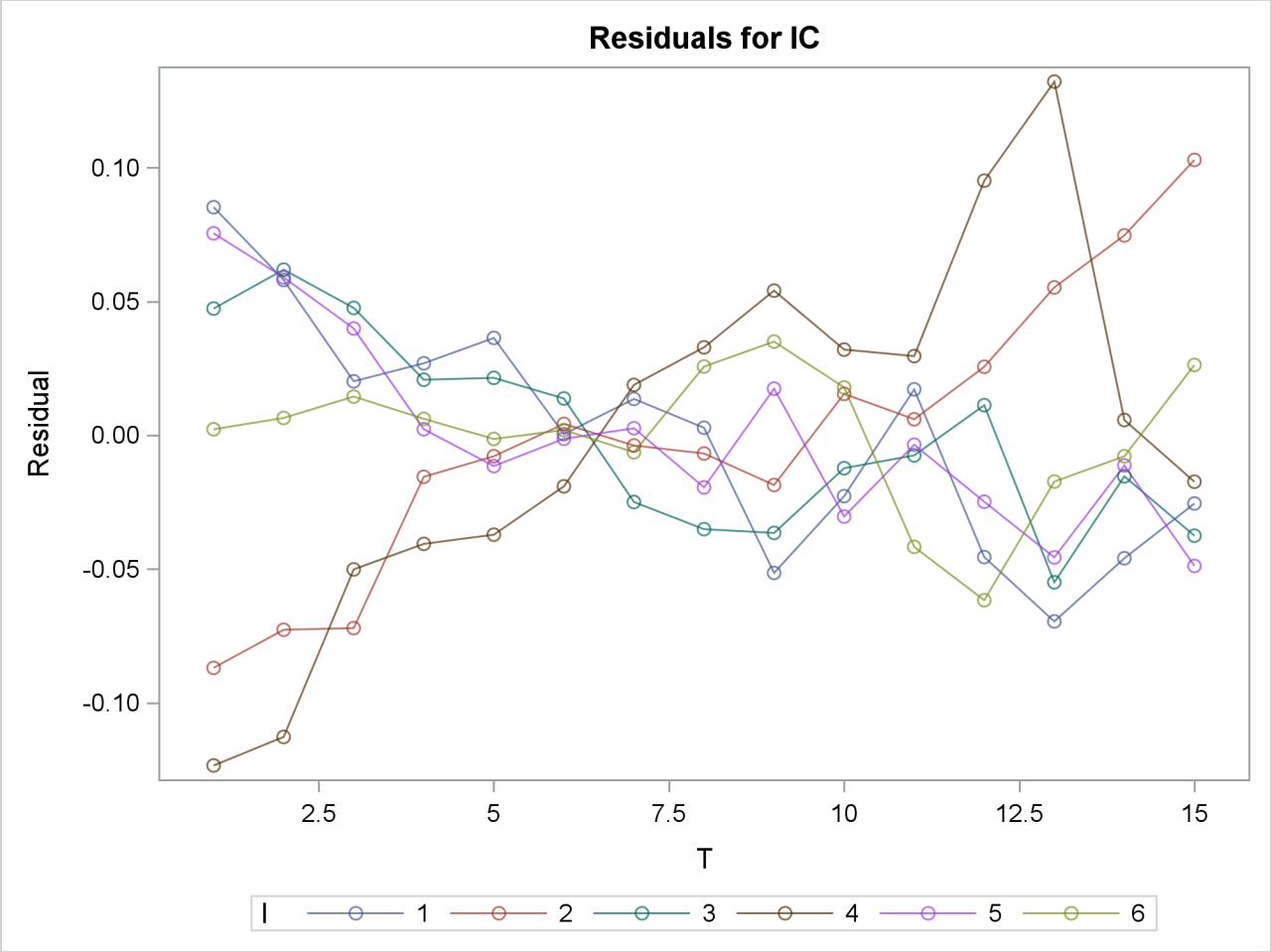
Output 26.2.6 Diagnostic Panel 2



The UNPACK and ONLY options produce individual detail images of paneled plots. The graph shown in Output 26.2.7 shows a detail plot of residuals by cross section. The packed version always puts all cross sections on one plot while the unpacked one shows the cross sections in groups of ten to avoid loss of detail.

```
proc panel data=airline;
   id i t;
   model lC = lQ lPF LF / fixtwo plots(unpack only) = residsurface;
run;
```

Output 26.2.7 Surface Plot of the Residual



Example 26.3: The Airline Cost Data: Further Analysis

Using the same data as in Example 26.2, you further investigate the "true" effect of fuel prices. Specifically, you run the FixOne model, ignoring time effects. You specify the following statements in PROC PANEL to run this model:

```
proc panel data=airline;
   id i t;
   model lC = lQ lPF LF / fixone;
run;
```

The preceding statements result in Output 26.3.1. The fit seems to have deteriorated somewhat. The SSE rises from 0.1768 to 0.2926.

Output 26.3.1 The Airline Cost Data—Fit Statistics

The PANEL Procedure
Fixed One-Way Estimates

Dependent Variable: lC (Log Transformation of Costs)

Fit Statistics			
SSE	0.2926	**DFE**	81
MSE	0.0036	**Root MSE**	0.0601
R-Square	0.9974		

You still reject poolability based on the F test in Output 26.3.2 at all accepted levels of significance.

Output 26.3.2 The Airline Cost Data—Test for Fixed Effects

F Test for No Fixed Effects			
Num DF	**Den DF**	**F Value**	**Pr > F**
5	81	57.74	<.0001

The parameters change somewhat dramatically as shown in Output 26.3.3. The effect of fuel costs comes in very strong and significant. The load factor's coefficient increases, although not as dramatically. This suggests that the fixed time effects might be proxies for both the oil shocks and deregulation.

Output 26.3.3 The Airline Cost Data—Parameter Estimates

Parameter Estimates						
Variable	**DF**	**Estimate**	**Standard Error**	**t Value**	**Pr > |t|**	**Label**
Intercept	1	9.79304	0.2636	37.15	<.0001	Intercept
lQ	1	0.919293	0.0299	30.76	<.0001	Log Transformation of Quantity
lPF	1	0.417492	0.0152	27.47	<.0001	Log Transformation of Price of Fuel
LF	1	-1.07044	0.2017	-5.31	<.0001	Load Factor (utilization index)

Example 26.4: The Airline Cost Data: Random-Effects Models

This example continues to use the Christenson Associates airline data, which measures costs, prices of inputs, and utilization rates for six airlines over the time span 1970–1984. There are six cross sections and fifteen time observations. Here, you examine the different estimates that are generated from the one-way random-effects and two-way random-effects models, by using four different methods to estimate the variance components: Fuller and Battese, Wansbeek and Kapteyn, Wallace and Hussain, and Nerlove.

The data for this example are created by the PROC PANEL statements shown in Example 26.2. The following PROC PANEL statements generate the estimates:

```
proc panel data=airline;
   id I T;
   model "One-Way, FB"    lC = lQ lPF lF / ranone vcomp=fb;
   model "One-Way, WK"    lC = lQ lPF lF / ranone vcomp=wk;
   model "One-Way, WH"    lC = lQ lPF lF / ranone vcomp=wh;
   model "One-Way, NL"    lC = lQ lPF lF / ranone vcomp=nl;
   model "Two-Way, FB"    lC = lQ lPF lF / rantwo vcomp=fb;
   model "Two-Way, WK"    lC = lQ lPF lF / rantwo vcomp=wk;
   model "Two-Way, WH"    lC = lQ lPF lF / rantwo vcomp=wh;
   model "Two-Way, NL"    lC = lQ lPF lF / rantwo vcomp=nl;
   model "Pooled"         lC = lQ lPF lF / pooled;
   model "Between Groups" lC = lQ lPF lF / btwng;
   model "Between Times"  lC = lQ lPF lF / btwnt;
   compare / pstat(estimate) mstat(varcs varts varerr);
run;
```

The parameter estimates and variance components for all models and estimators are reported in Output 26.4.1 and Output 26.4.2. Both tables are created by the COMPARE statement.

Output 26.4.1 Parameter Estimates

The PANEL Procedure
Model Comparison

Dependent Variable: lC (Log Transformation of Costs)

Comparison of Model Parameter Estimates							
Variable		One-Way, FB RanOne	One-Way, WK RanOne	One-Way, WH RanOne	One-Way, NL RanOne	Two-Way, FB RanTwo	Two-Way, WK RanTwo
Intercept	Estimate	9.637027	9.629542	9.643869	9.640560	9.362705	9.643579
lQ	Estimate	0.908032	0.906926	0.909042	0.908554	0.866458	0.843341
lPF	Estimate	0.422199	0.422676	0.421766	0.421975	0.436160	0.409662
LF	Estimate	-1.064733	-1.064564	-1.064966	-1.064844	-0.980482	-0.926308

Comparison of Model Parameter Estimates						
Variable		Two-Way, WH RanTwo	Two-Way, NL RanTwo	Pooled Pooled	Between Groups BtwGrps	Between Times BtwTime
Intercept	Estimate	9.379328	9.972603	9.516907	85.809402	11.184905
lQ	Estimate	0.869214	0.838724	0.882740	0.782455	1.133318
lPF	Estimate	0.435317	0.382904	0.453978	-5.524011	0.334268
LF	Estimate	-0.985181	-0.913357	-1.627511	-1.750949	-1.350947

Output 26.4.2 Variance Component Estimates

The PANEL Procedure
Model Comparison

Dependent Variable: lC (Log Transformation of Costs)

Comparison of Model Statistics							
Statistic	**One-Way, FB RanOne**	**One-Way, WK RanOne**	**One-Way, WH RanOne**	**One-Way, NL RanOne**	**Two-Way, FB RanTwo**	**Two-Way, WK RanTwo**	**Two-Way, WH RanTwo**
Var due to Cross Sections	0.0182	0.0160	0.0187	0.0174	0.0174	0.0156	0.0187
Var due to Time Series					0.001081	0.0391	0.000854
Var due to Error	0.003612	0.003612	0.003280	0.003251	0.002639	0.002639	0.002502

Comparison of Model Statistics				
Statistic	**Two-Way, NL RanTwo**	**Pooled Pooled**	**Between Groups BtwGrps**	**Between Times BtwTime**
Var due to Cross Sections	0.0171			
Var due to Time Series	0.0591			
Var due to Error	0.001965	0.0155	0.0158	0.000508

In the random-effects model, individual constant terms are viewed as randomly distributed across cross-sectional units. They are not viewed as parametric shifts of the regression function, as they are in the fixed-effects model. This is appropriate when the sampled cross-sectional units are drawn from a large population. In this example, the six airlines are clearly a sample of all the airlines in the industry and not an exhaustive, or nearly exhaustive, list.

There are four techniques for computing the variance components in the one-way random-effects model. The method by Fuller and Battese (1974) (FB) uses a "fitting of constants" method to estimate them, the Wansbeek and Kapteyn (1989) (WK) method uses the true disturbances, the Wallace and Hussain (1969) (WH) method uses ordinary least squares residuals, and the Nerlove (1971) (NL) method uses one-way fixed-effects regression.

Looking at the estimates of the variance components for cross section and error in Output 26.4.2, you see that equal variance components for error for FB and WK are equal, whereas they are nearly equal for WH and NL.

All four techniques produce different variance components for cross sections. These estimates are then used to estimate the values of the parameters in Output 26.4.1. All the parameters appear to have similar and equally plausible estimates. Both the indices for output in revenue passenger miles (lQ) and fuel price (lPF) have small, positive effects on total costs, which you would expect. The load factor (LF) has a somewhat larger and negative effect on total costs, suggesting that as utilization increases, costs decrease.

Comparing the four two-way estimators, you find that the variance components for error produced by the FB and WK methods are equal, as they are in the one-way model. However, in this case, the WH and NL methods produce variance estimates that are dissimilar. The estimates of the variance component for cross sections are all different, but in a close range. The same cannot be said for the variance component for time series. As varied as each of the variance estimates might be, they produce parameter estimates that are similar and plausible. As in the one-way effects model, the indices for output (lQ) and fuel price (lPF) are

small and positive. The load factor (LF) estimates are all negative and somewhat smaller in magnitude than the estimates that are produced in the one-way model. During the time the data were collected, the Civil Aeronautics Board dissolved, so it is possible that the time dummy variables are proxies for this dissolution. The dissolution would lead to the decay of time effects and an imprecise estimation of the effects of the load factors, even though the estimates are statistically significant.

The pooled estimates give you something to compare the random-effects estimates against. You see that signs and magnitudes of output and fuel price are similar but that the magnitude of the load factor coefficient is somewhat larger under pooling. Because the model appears to have both cross-sectional and time series effects, the pooled model should not be used.

Finally, you examine the between estimators. For the between-groups estimate, you are looking at each airline's data averaged across time. You see in Output 26.4.1 that the between-groups parameter estimates are radically different from all other parameter estimates. This difference could indicate that the time component is not being appropriately handled with this technique. For the between-times estimate, you are looking at the average across all airlines in each time period. In this case, the parameter estimates are of the same sign and closer in magnitude to the previously computed estimates. Both the output and load factor effects appear to have more bearing on total costs.

Example 26.5: Panel Study of Income Dynamics (PSID): Hausman-Taylor Models

Cornwell and Rupert (1988) analyze data from the Panel Study of Income Dynamics (PSID), an income study of 595 individuals over the seven-year period, 1976–1982 inclusive. Of particular interest is the effect of additional schooling on wages. The analysis here replicates that of Baltagi (2008, sec. 7.5), where it is surmised that covariate correlation with individual effects makes a standard random-effects model inadequate.

The following statements create the PSID data:

```
data psid;
   input id t lwage wks south smsa ms exp exp2 occ ind union fem blk ed;
   label id    = 'Person ID'
         t     = 'Time'
         lwage = 'Log(wages)'
         wks   = 'Weeks worked'
         south = '1 if resides in the South'
         smsa  = '1 if resides in SMSA'
         ms    = '1 if married'
         exp   = 'Years full-time experience'
         exp2  = 'exp squared'
         occ   = '1 if blue-collar occupation'
         ind   = '1 if manufacturing'
         union = '1 if union contract'
         fem   = '1 if female'
         blk   = '1 if black'
         ed    = 'Years of education';
datalines;
1    1  5.5606799126  32  1  0  1  3   9     0  0  0  0  0  9
1    2  5.7203102112  43  1  0  1  4   16    0  0  0  0  0  9
1    3  5.9964499474  40  1  0  1  5   25    0  0  0  0  0  9
```

```
1    4  5.9964499474  39  1  0  1  6   36    0  0  0  0  0  9
1    5  6.0614600182  42  1  0  1  7   49    0  1  0  0  0  9
1    6  6.1737899780  35  1  0  1  8   64    0  1  0  0  0  9
1    7  6.2441701889  32  1  0  1  9   81    0  1  0  0  0  9
2    1  6.1633100510  34  0  0  1  30  900   1  0  0  0  0  11
2    2  6.2146100998  27  0  0  1  31  961   1  0  0  0  0  11
2    3  6.2634000778  33  0  0  1  32  1024  1  1  1  0  0  11
2    4  6.5439100266  30  0  0  1  33  1089  1  1  0  0  0  11
2    5  6.6970300674  30  0  0  1  34  1156  1  1  0  0  0  11
2    6  6.7912201881  37  0  0  1  35  1225  1  1  0  0  0  11
2    7  6.8156399727  30  0  0  1  36  1296  1  1  0  0  0  11

   ... more lines ...
```

You begin by fitting a one-way random effects model:

```
proc panel data=psid;
   id id t;
   model lwage = wks south smsa ms exp exp2 occ
                          ind union fem blk ed / ranone;
run;
```

The output is shown in Output 26.5.1. The coefficient on the variable ED (which represents years of education) estimates that an additional year of schooling is associated with about a 10.7% increase in wages. However, the results of the Hausman test for random effects show a serious violation of the random-effects assumptions, namely that the regressors are independent of both error components.

Output 26.5.1 One-Way Random Effects Estimation

The PANEL Procedure
Fuller and Battese Variance Components (RanOne)

Dependent Variable: lwage (Log(wages))

Model Description	
Estimation Method	RanOne
Number of Cross Sections	595
Time Series Length	7

Variance Component Estimates	
Variance Component for Cross Sections	0.100553
Variance Component for Error	0.023102

Hausman Test for Random Effects			
Coefficients	DF	m Value	Pr > m
9	9	5288.98	<.0001

Output 26.5.1 *continued*

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Pr > |t|	Label
Intercept	1	4.030811	0.1044	38.59	<.0001	Intercept
wks	1	0.000954	0.000740	1.29	0.1971	Weeks worked
south	1	-0.00788	0.0281	-0.28	0.7795	1 if resides in the South
smsa	1	-0.02898	0.0202	-1.43	0.1517	1 if resides in SMSA
ms	1	-0.07067	0.0224	-3.16	0.0016	1 if married
exp	1	0.087726	0.00281	31.27	<.0001	Years full-time experience
exp2	1	-0.00076	0.000062	-12.31	<.0001	exp squared
occ	1	-0.04293	0.0162	-2.65	0.0081	1 if blue-collar occupation
ind	1	0.00381	0.0172	0.22	0.8242	1 if manufacturing
union	1	0.058121	0.0169	3.45	0.0006	1 if union contract
fem	1	-0.30791	0.0572	-5.38	<.0001	1 if female
blk	1	-0.21995	0.0660	-3.33	0.0009	1 if black
ed	1	0.10742	0.00642	16.73	<.0001	Years of education

An alternative could be a fixed-effects (FIXONE) model, but that would not permit estimation of the coefficient for ED, which does not vary within individuals. A compromise is the Hausman-Taylor model, for which you stipulate a set of covariates that are correlated with the individual effects (but uncorrelated with the observation-level errors). You specify the correlated variables in the CORRELATED= option in the INSTRUMENTS statement:

```
proc panel data=psid;
   id id t;
   instruments correlated = (wks ms exp exp2 union ed);
   model lwage = wks south smsa ms exp exp2 occ
                            ind union fem blk ed / htaylor;
run;
```

The results are shown in Output 26.5.2. The table of parameter estimates has an added column, Type, that identifies which regressors are assumed to be correlated with individual effects (C) and which regressors do not vary within cross sections (TI). It was stated previously that the Hausman-Taylor model is a compromise between fixed-effects and random-effects models, and you can think of the compromise this way: You want to fit a random-effects model, but the correlated (C) variables make that model invalid. So you fall back to the consistent fixed-effects model, but then the time-invariant (TI) variables are the problem because they will be dropped from that model. The solution is to use the Hausman-Taylor estimator.

The estimation results show that an additional year of schooling is now associated with a 13.8% increase in wages. Also presented is a Hausman test that compares this model to the fixed-effects model. As was the case previously when you fit the random-effects model, you can think of the Hausman test as a referendum on the assumptions you are making. For this estimation, it seems that your choice of variables to treat as correlated is adequate. It also seems to hold true that any correlation is with the individual-level effects, and not the observation-level errors.

Output 26.5.2 Hausman-Taylor Estimation

The PANEL Procedure
Hausman and Taylor Model for Correlated Individual Effects (HTaylor)

Dependent Variable: lwage (Log(wages))

Variance Component Estimates	
Variance Component for Cross Sections	0.886993
Variance Component for Error	0.023044

Hausman Test against Fixed Effects			
Coefficients	DF	m Value	Pr > m
9	3	5.26	0.1539

Parameter Estimates

Variable	Type		DF	Estimate	Standard Error	t Value	Pr > |t|	Label
Intercept			1	2.912726	0.2837	10.27	<.0001	Intercept
wks	C		1	0.000837	0.000600	1.40	0.1627	Weeks worked
south			1	0.00744	0.0320	0.23	0.8159	1 if resides in the South
smsa			1	-0.04183	0.0190	-2.21	0.0274	1 if resides in SMSA
ms	C		1	-0.02985	0.0190	-1.57	0.1159	1 if married
exp	C		1	0.113133	0.00247	45.79	<.0001	Years full-time experience
exp2	C		1	-0.00042	0.000055	-7.67	<.0001	exp squared
occ			1	-0.0207	0.0138	-1.50	0.1331	1 if blue-collar occupation
ind			1	0.013604	0.0152	0.89	0.3720	1 if manufacturing
union	C		1	0.032771	0.0149	2.20	0.0280	1 if union contract
fem		TI	1	-0.13092	0.1267	-1.03	0.3014	1 if female
blk		TI	1	-0.28575	0.1557	-1.84	0.0665	1 if black
ed	C	TI	1	0.137944	0.0212	6.49	<.0001	Years of education

C: correlated with the individual effects
TI: constant (time-invariant) within cross sections

At its core, the Hausman-Taylor estimator is an instrumental variables regression, where the instruments are derived from regressors that are assumed to be uncorrelated with the individual effects. Technically it is the cross-sectional means of these variables that need to be uncorrelated, not the variables themselves.

The Amemiya-MaCurdy model is a close relative of the Hausman-Taylor model. The only difference between the two is that the Amemiya-MaCurdy model makes the added assumption that the regressors (and not just their means) are uncorrelated with the individual effects. By making that assumption, the Amemiya-MaCurdy model can take advantage of a more efficient set of instrumental variables.

The following statements fit the Amemiya-MaCurdy model:

```
proc panel data=psid;
   id id t;
   instruments correlated = (wks ms exp exp2 union ed);
   model lwage = wks south smsa ms exp exp2 occ
                          ind union fem blk ed / amacurdy;
run;
```

The results are shown in Output 26.5.3. Little is changed from the Hausman-Taylor model. The Hausman test compares the Amemiya-MaCurdy model to the Hausman-Taylor model (not the fixed-effects model as previously) and shows that the one additional assumption is acceptable. You even gained a bit of efficiency in the process; compare the standard deviations of the coefficient on the variable ED from both models.

Output 26.5.3 Amemiya-MaCurdy Estimation

The PANEL Procedure
Amemiya and MaCurdy Model for Correlated Individual Effects (AMaCurdy)

Dependent Variable: lwage (Log(wages))

Variance Component Estimates	
Variance Component for Cross Sections	0.886993
Variance Component for Error	0.023044

Hausman Test against Hausman-Taylor			
Coefficients	DF	m Value	Pr > m
13	13	14.67	0.3287

Parameter Estimates

Variable	Type		DF	Estimate	Standard Error	t Value	Pr > |t|	Label
Intercept			1	2.927338	0.2751	10.64	<.0001	Intercept
wks	C		1	0.000838	0.000599	1.40	0.1622	Weeks worked
south			1	0.007282	0.0319	0.23	0.8197	1 if resides in the South
smsa			1	-0.04195	0.0189	-2.21	0.0269	1 if resides in SMSA
ms	C		1	-0.03009	0.0190	-1.59	0.1127	1 if married
exp	C		1	0.11297	0.00247	45.76	<.0001	Years full-time experience
exp2	C		1	-0.00042	0.000055	-7.72	<.0001	exp squared
occ			1	-0.02085	0.0138	-1.51	0.1299	1 if blue-collar occupation
ind			1	0.013629	0.0152	0.89	0.3709	1 if manufacturing
union	C		1	0.032475	0.0149	2.18	0.0293	1 if union contract
fem		TI	1	-0.13201	0.1266	-1.04	0.2972	1 if female
blk		TI	1	-0.2859	0.1555	-1.84	0.0660	1 if black
ed	C	TI	1	0.137205	0.0206	6.67	<.0001	Years of education

C: correlated with the individual effects
TI: constant (time-invariant) within cross sections

Finally, you should realize that the Hausman-Taylor and Amemiya-MaCurdy estimators are not cure-alls for correlated individual effects. Estimation tacitly relies on the uncorrelated regressors being sufficient to predict the correlated regressors. Otherwise, you run into the problem of weak instruments. If you have weak instruments, you will obtain biased estimates that have very large standard errors. However, that does not seem to be the case here.

Example 26.6: Dynamic Panel Estimation of Cigarette Sales Data

In this example, a dynamic panel demand model for cigarette sales is estimated. It illustrates the application of the method described in the section "Dynamic Panel Estimators" on page 1852. The data are a panel from

46 American states over the period 1963–1992. For a description of the data, see Baltagi and Levin (1992); (Baltagi 2008, sec. 8.8). All variables were transformed by taking the natural logarithm. The data set CIGAR is shown in the following statements:

```
data cigar;
   input state year price pop pop_16 cpi ndi sales pimin;
   label
   state  = 'State abbreviation'
   year   = 'YEAR'
   price  = 'Price per pack of cigarettes'
   pop    = 'Population'
   pop_16 = 'Population above the age of 16'
   cpi    = 'Consumer price index with (1983=100)'
   ndi    = 'Per capita disposable income'
   sales  = 'Cigarette sales in packs per capita'
   pimin  = 'Minimum price in adjoining states per pack of cigarettes';
datalines;
1 63 28.6 3383 2236.5 30.6 1558.3045298 93.9 26.1
1 64 29.8 3431 2276.7 31.0 1684.0732025 95.4 27.5
1 65 29.8 3486 2327.5 31.5 1809.8418752 98.5 28.9
1 66 31.5 3524 2369.7 32.4 1915.1603572 96.4 29.5
1 67 31.6 3533 2393.7 33.4 2023.5463678 95.5 29.6
1 68 35.6 3522 2405.2 34.8 2202.4855362 88.4 32
1 69 36.6 3531 2411.9 36.7 2377.3346665 90.1 32.8
1 70 39.6 3444 2394.6 38.8 2591.0391591 89.8 34.3
1 71 42.7 3481 2443.5 40.5 2785.3159706 95.4 35.8
1 72 42.3 3511 2484.7 41.8 3034.8082969 101.1 37.4

   ... more lines ...
```

The following statements sort the data by STATE and YEAR variables:

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

Next, logarithms of the variables required for regression estimation are calculated, as shown in the following statements:

```
data cigar;
   set cigar;
   lsales = log(sales);
   lprice = log(price);
   lndi = log(ndi);
   lpimin = log(pimin);
   label lprice = 'Log price per pack of cigarettes';
   label lndi = 'Log per capita disposable income';
   label lsales = 'Log cigarette sales in packs per capita';
   label lpimin = 'Log minimum price in adjoining states
                     per pack of cigarettes';
run;
```

The following statements create the CIGAR_LAG data set with lagged variable for each cross section:

```
proc panel data=cigar;
   id state year;
   clag lsales(1) / out=cigar_lag;
run;

data cigar_lag;
   set cigar_lag;
   label lsales_1 = 'Lagged log cigarette sales in packs per capita';
run;
```

Finally, the model is estimated by a two step GMM method. Five lags (MAXBAND=5) of the dependent variable are used as instruments. The NOLEVELS option is specified to avoid the use of level equations.

```
proc panel data=cigar_lag;
   inst depvar;
   model lsales = lsales_1 lprice lndi lpimin
         / gmm2 nolevels maxband=5 noint;
   id state year;
run;
```

Output 26.6.1 Estimation with Two-step GMM

The PANEL Procedure
GMM: First Differences Transformation

Dependent Variable: lsales (Log cigarette sales in packs per capita)

Model Description	
Estimation Method	GMM2
Number of Cross Sections	46
Time Series Length	30
Estimate Stage	2
Maximum Number of Time Periods (MAXBAND)	5

Fit Statistics			
SSE	2187.5988	DFE	1284
MSE	1.7037	Root MSE	1.3053

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Pr > |t|
lsales_1	1	0.572219	0.000981	583.51	<.0001
lprice	1	-0.23464	0.00306	-76.56	<.0001
lndi	1	0.232673	0.000392	593.69	<.0001
lpimin	1	-0.08299	0.00328	-25.29	<.0001

If the theory suggests that there are other valid instruments, the PREDETERMINED, EXOGENOUS and CORRELATED options can also be used.

Example 26.7: Using the FLATDATA Statement

Sometimes data sets are stored in compressed (or wide) form, where each record contains all observations for the entire cross section. Although the PANEL procedure requires data in uncompressed (long) form, sometimes it is easier to create new variables or summary statistics if the data are in wide form.

To illustrate, suppose you have a simulated data set that contains 20 cross sections measured over six time periods. Each time period has values for dependent and independent variables, $Y_1, \ldots, Y_6$ and $X_1, \ldots, X_6$. The cs and num variables are constant across each cross section.

The observations for the first five cross sections along with other variables are shown in Output 26.7.1. In this example, *i* represents the cross section. The time period is identified by the subscript of the Y and X variables; it ranges from 1 to 6.

Output 26.7.1 Compressed Data Set

Obs	i	cs	num	X_1	X_2	X_3	X_4	X_5	X_6	Y_1	Y_2
1	1	CS1	-1.56058	0.40268	0.91951	0.69482	-2.28899	-1.32762	1.92348	2.30418	2.11850
2	2	CS2	0.30989	1.01950	-0.04699	-0.96695	-1.08345	-0.05180	0.30266	4.50982	3.73887
3	3	CS3	0.85054	0.60325	0.71154	0.66168	-0.66823	-1.87550	0.55065	4.07276	4.89621
4	4	CS4	-0.18885	-0.64946	-1.23355	0.04554	-0.24996	0.09685	-0.92771	2.40304	1.48182
5	5	CS5	-0.04761	-0.79692	0.63445	-2.23539	-0.37629	-0.82212	-0.70566	3.58092	6.08917

Obs	Y_3	Y_4	Y_5	Y_6
1	2.66009	-4.94104	-0.83053	5.01359
2	1.44984	-1.02996	2.78260	1.73856
3	3.90470	1.03437	0.54598	5.01460
4	2.70579	3.82672	4.01117	1.97639
5	3.08249	4.26605	3.65452	0.81826

When the data are in this form, it is easy to create other variables that are combinations of the existing variables. For example, you can calculate the within-cross-section mean of *X* by simply summing across the *X_i* variables and dividing by six. It is easier to perform this kind of data manipulation when the data are in compressed (wide) form instead of uncompressed (long) form.

On the other hand, the PANEL procedure cannot work directly with the data in wide form. You can use the FLATDATA statement to transform wide data into long form "on the fly" for performing a panel-data analysis. You can also use the OUT= option to output the transformed data to a new data set, to use for further analysis.

The following code reshapes the data and performs fixed-effects estimation:

```
proc panel data=flattest;
   flatdata indid=i tsname="t" base=(X Y)
            keep=( cs num seed )  / out=flat_out;
   id i t;
   model y = x / fixone noint;
run;
```

The first six observations for the uncompressed (long) data set and the results for the one-way fixed-effects model are shown in Output 26.7.2 and Output 26.7.3, respectively.

Output 26.7.2 Uncompressed Data Set

Obs	I	t	X	Y	CS	NUM
1	1	1	0.40268	2.30418	CS1	-1.56058
2	1	2	0.91951	2.11850	CS1	-1.56058
3	1	3	0.69482	2.66009	CS1	-1.56058
4	1	4	-2.28899	-4.94104	CS1	-1.56058
5	1	5	-1.32762	-0.83053	CS1	-1.56058
6	1	6	1.92348	5.01359	CS1	-1.56058

Output 26.7.3 Estimation with the FLATDATA Statement

The PANEL Procedure
Fixed One-Way Estimates

Dependent Variable: Y

Parameter Estimates

Variable	DF	Estimate	Standard Error	t Value	Pr > |t|	Label
X	1	2.010753	0.1217	16.52	<.0001	

Now, suppose you have long data that you want to reshape into wide form. The following DATA step performs this task:

```
data wide;
   set flat_out;
   by i;
   keep i num cs X_1-X_6 Y_1-Y_6;
   retain X_1-X_6 Y_1-Y_6;
   array ax(1:6) X_1-X_6;
   array ay(1:6) Y_1-Y_6;
   if first.i then do;
      do j = 1 to 6;
         ax(j) = 0;
         ay(j) = 0;
      end;
   end;
   ax(t) = X;
   ay(t) = Y;
   if last.i then output;
run;
```

As a check, Output 26.7.4 lists the newly compressed data, which match the original data from this example.

Output 26.7.4 Recompressed Data Set

Obs	I	CS	NUM	X_1	X_2	X_3	X_4	X_5	X_6	Y_1	Y_2
1	1	CS1	-1.56058	0.40268	0.91951	0.69482	-2.28899	-1.32762	1.92348	2.30418	2.11850
2	2	CS2	0.30989	1.01950	-0.04699	-0.96695	-1.08345	-0.05180	0.30266	4.50982	3.73887
3	3	CS3	0.85054	0.60325	0.71154	0.66168	-0.66823	-1.87550	0.55065	4.07276	4.89621
4	4	CS4	-0.18885	-0.64946	-1.23355	0.04554	-0.24996	0.09685	-0.92771	2.40304	1.48182
5	5	CS5	-0.04761	-0.79692	0.63445	-2.23539	-0.37629	-0.82212	-0.70566	3.58092	6.08917

Obs	Y_3	Y_4	Y_5	Y_6
1	2.66009	-4.94104	-0.83053	5.01359
2	1.44984	-1.02996	2.78260	1.73856
3	3.90470	1.03437	0.54598	5.01460
4	2.70579	3.82672	4.01117	1.97639
5	3.08249	4.26605	3.65452	0.81826

References

Amemiya, T., and MaCurdy, T. E. (1986). "Instrumental-Variable Estimation of an Error-Components Model." *Econometrica* 54:869–880.

Andrews, D. W. K. (1991). "Heteroscedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica* 59:817–858.

Arellano, M. (1987). "Computing Robust Standard Errors for Within-Groups Estimators." *Oxford Bulletin of Economics and Statistics* 49:431–434.

Arellano, M., and Bond, S. (1991). "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Review of Economic Studies* 58:277–297.

Arellano, M., and Bover, O. (1995). "Another Look at the Instrumental Variable Estimation of Error-Components Models." *Journal of Econometrics* 68:29–51.

Baltagi, B. H. (2008). *Econometric Analysis of Panel Data*. 4th ed. Chichester, UK: John Wiley & Sons.

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* 62:67–89.

Baltagi, B. H., Chang, Y. J., and Li, Q. (1992). "Monte Carlo Results on Several New and Existing Tests for the Error Component Model." *Journal of Econometrics* 54:95–120.

Baltagi, B. H., and Levin, D. (1992). "Cigarette Taxation: Raising Revenues and Reducing Consumption." *Structural Change and Economic Dynamics* 3:321–335.

Baltagi, B. H., and Li, Q. (1991). "A Joint Test for Serial Correlation and Random Individual Effects." *Statistics and Probability Letters* 11:277–280.

Baltagi, B. H., and Li, Q. (1995). "Testing AR(1) against MA(1) Disturbances in an Error Component Model." *Journal of Econometrics* 68:133–151.

Baltagi, B. H., Song, S. H., and Jung, B. C. (2002). "A Comparative Study of Alternative Estimators for the Unbalanced Two-Way Error Component Regression Model." *Econometrics Journal* 5:480–493.

Bera, A. K., Sosa Escudero, W., and Yoon, M. (2001). "Tests for the Error Component Model in the Presence of Local Misspecification." *Journal of Econometrics* 101:1–23.

Bhargava, A., Franzini, L., and Narendranathan, W. (1982). "Serial Correlation and Fixed Effects Model." *Review of Economic Studies* 49:533–549.

Blundell, R., and Bond, S. (1998). "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models." *Journal of Econometrics* 87:115–143.

Breitung, J. (2000). "The Local Power of Some Unit Root Tests for Panel Data." In *Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, edited by B. H. Baltagi, 161–178. Vol. 15 of Advances in Econometrics. Amsterdam: JAI Press.

Breitung, J., and Das, S. (2005). "Panel Unit Root Tests under Cross-Sectional Dependence." *Statistica Neerlandica* 59:414–433.

Breitung, J., and Meyer, W. (1994). "Testing for Unit Roots in Panel Data: Are Wages on Different Bargaining Levels Cointegrated?" *Applied Economics* 26:353–361.

Breusch, T. S., and Pagan, A. R. (1980). "The Lagrange Multiplier Test and Its Applications to Model Specification in Econometrics." *Review of Econometric Studies* 47:239–253.

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

Campbell, J. Y., and Perron, P. (1991). "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots." In *NBER Macroeconomics Annual*, edited by O. Blanchard and S. Fisher, 141–201. Cambridge, MA: MIT Press.

Choi, I. (2001). "Unit Root Tests for Panel Data." *Journal of International Money and Finance* 20:249–272.

Choi, I. (2006). "Nonstationary Panels." In *Econometric Theory*, edited by T. C. Mills and K. Patterson, 511–539. Vol. 1 of Palgrave Handbook of Econometrics. Basingstoke, UK: Palgrave Macmillan.

Chow, G. (1960). "Tests of Equality between Sets of Coefficients in Two Linear Regressions." *Econometrica* 28:531–534.

Cornwell, C., and Rupert, P. (1988). "Efficient Estimation with Panel Data: An Empirical Comparison of Instrumental Variables Estimators." *Journal of Applied Econometrics* 3:149–155.

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

Davidson, R., and MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics*. New York: Oxford University Press.

Davis, P. (2002). "Estimating Multi-way Error Components Models with Unbalanced Data Structures." *Journal of Econometrics* 106:67–95.

Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996). "Efficient Tests for an Autoregressive Unit Root." *Econometrica* 64:813–836.

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

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

Fisher, R. A. (1932). *Statistical Methods for Research Workers*. 4th ed. Edinburgh: Oliver & Boyd.

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

Gardner, R. (1998). "Unobservable Individual Effects in Unbalanced Panel Data." *Economics Letters* 58:39–42.

Gourieroux, C., Holly, A., and Monfort, A. (1982). "Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters." *Econometrica* 50:63–80.

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

Greene, W. H. (2000). *Econometric Analysis*. 4th ed. Upper Saddle River, NJ: Prentice-Hall.

Hadri, K. (2000). "Testing for Stationarity in Heterogeneous Panel Data." *Econometrics Journal* 3:148–161.

Hall, A. R. (1994). "Testing for a Unit Root with Pretest Data Based Model Selection." *Journal of Business and Economic Statistics* 12:461–470.

Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hannan, E. J., and Quinn, B. G. (1979). "The Determination of the Order of an Autoregression." *Journal of the Royal Statistical Society, Series B* 41:190–195.

Harris, R. D. F., and Tzavalis, E. (1999). "Inference for Unit Roots in Dynamic Panels Where the Time Dimension Is Fixed." *Journal of Econometrics* 91:201–226.

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

Hausman, J. A., and Taylor, W. E. (1981). "Panel Data and Unobservable Individual Effects." *Econometrica* 49:1377–1398.

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

Honda, Y. (1985). "Testing the Error Components Model with Non-normal Disturbances." *Review of Economic Studies* 52:681–690.

Honda, Y. (1991). "A Standardized Test for the Error Components Model with the Two-Way Layout." *Economics Letters* 37:125–128.

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

Im, K. S., Pesaran, M. H., and Shin, Y. (2003). "Testing for Unit Root in Heterogeneous Panels." *Journal of Econometrics* 115:53–74.

Judge, G. G., Griffiths, W. E., Hill, R. C., Lütkepohl, H., and Lee, T.-C. (1985). *The Theory and Practice of Econometrics*. 2nd ed. New York: John Wiley & Sons.

King, M. L., and Wu, P. X. (1997). "Locally Optimal One-Sided Tests for Multiparameter Hypotheses." *Econometric Reviews* 16:131–156.

Kmenta, J. (1971). *Elements of Econometrics*. New York: Macmillan.

LaMotte, L. R. (1994). "A Note on the Role of Independence in *t* Statistics Constructed from Linear Statistics in Regression Models." *American Statistician* 48:238–240.

Levin, A., Lin, C.-F., and Chu, C. S. (2002). "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties." *Journal of Econometrics* 108:1–24.

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

Maddala, G. S., and Wu, S. (1999). "A Comparative Study of Unit Root Tests with Panel Data and a New Simple Test." *Oxford Bulletin of Economics and Statistics* 61:631–652.

Moulton, B. R., and Randolph, W. C. (1989). "Alternative Tests of the Error Components Model." *Econometrica* 57:685–693.

Nabeya, S. (1999). "Asymptotic Moments of Some Unit Root Test Statistics in the Null Case." *Econometric Theory* 15:139–149.

Nerlove, M. (1971). "Further Evidence on the Estimation of Dynamic Relations from a Time Series of Cross Sections." *Econometrica* 39:359–382.

Newey, W. K., and West, D. W. (1994). "Automatic Lag Selection in Covariance Matrix Estimation." *Review of Economic Studies* 61:631–653.

Ng, S., and Perron, P. (2001). "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power." *Econometrica* 69:1519–1554.

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.

Pesaran, M. H. (2004). "General Diagnostic Tests for Cross Section Dependence in Panels." Institute for the Study of Labor (IZA) Discussion Paper No. 1240, CESifo Working Paper No. 1229, Department of Applied Economics, University of Cambridge.

Phillips, P. C. B., and Perron, P. (1988). "Testing for a Unit Root in Time Series Regression." *Biometrika* 75:335–346.

Roy, S. N. (1957). *Some Aspects of Multivariate Analysis*. New York: John Wiley & Sons.

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. diss., 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 State University.

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

Stock, J. H., and Watson, M. W. (2002). *Introduction to Econometrics*. 3rd ed. Reading, MA: Addison-Wesley.

Theil, H. (1961). *Economic Forecasts and Policy*. 2nd ed. Amsterdam: North-Holland.

Wallace, T., and Hussain, A. (1969). "The Use of Error Components Model in Combining Cross Section with Time Series Data." *Econometrica* 37:55–72.

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

White, H. (1980). "A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48:817–838.

Windmeijer, F. (2005). "A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators." *Journal of Econometrics* 126:25–51.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

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

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.

Subject Index

Syntax Index