Gain Greater Insight into Your SAS® Software with SAS Books.

Discover all that you need on your journey to knowledge and empowerment.

support.sas.com/bookstore for additional books and resources.
# Chapter 20
## The PANEL Procedure

## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview: PANEL Procedure</td>
<td>1376</td>
</tr>
<tr>
<td>Getting Started: PANEL Procedure</td>
<td>1378</td>
</tr>
<tr>
<td>Syntax: PANEL Procedure</td>
<td>1380</td>
</tr>
<tr>
<td>Functional Summary</td>
<td>1380</td>
</tr>
<tr>
<td>PROC PANEL Statement</td>
<td>1383</td>
</tr>
<tr>
<td>BY Statement</td>
<td>1384</td>
</tr>
<tr>
<td>CLASS Statement</td>
<td>1385</td>
</tr>
<tr>
<td>FLATDATA Statement</td>
<td>1385</td>
</tr>
<tr>
<td>ID Statement</td>
<td>1386</td>
</tr>
<tr>
<td>INSTRUMENTS Statement</td>
<td>1386</td>
</tr>
<tr>
<td>LAG, ZLAG, XLAG, SLAG, or CLAG Statement</td>
<td>1388</td>
</tr>
<tr>
<td>MODEL Statement</td>
<td>1389</td>
</tr>
<tr>
<td>OUTPUT Statement</td>
<td>1401</td>
</tr>
<tr>
<td>RESTRICT Statement</td>
<td>1401</td>
</tr>
<tr>
<td>TEST Statement</td>
<td>1402</td>
</tr>
<tr>
<td>Details: PANEL Procedure</td>
<td>1403</td>
</tr>
<tr>
<td>Specifying the Input Data</td>
<td>1403</td>
</tr>
<tr>
<td>Specifying the Regression Model</td>
<td>1404</td>
</tr>
<tr>
<td>Unbalanced Data</td>
<td>1404</td>
</tr>
<tr>
<td>Missing Values</td>
<td>1405</td>
</tr>
<tr>
<td>Computational Resources</td>
<td>1405</td>
</tr>
<tr>
<td>Restricted Estimates</td>
<td>1405</td>
</tr>
<tr>
<td>Notation</td>
<td>1406</td>
</tr>
<tr>
<td>One-Way Fixed-Effects Model</td>
<td>1407</td>
</tr>
<tr>
<td>Two-Way Fixed-Effects Model</td>
<td>1408</td>
</tr>
<tr>
<td>Balanced Panels</td>
<td>1409</td>
</tr>
<tr>
<td>Unbalanced Panels</td>
<td>1411</td>
</tr>
<tr>
<td>First-Differenced Methods for One-Way and Two-Way Models</td>
<td>1414</td>
</tr>
<tr>
<td>Between Estimators</td>
<td>1414</td>
</tr>
<tr>
<td>Pooled Estimator</td>
<td>1415</td>
</tr>
<tr>
<td>One-Way Random-Effects Model</td>
<td>1415</td>
</tr>
<tr>
<td>Two-Way Random-Effects Model</td>
<td>1418</td>
</tr>
<tr>
<td>Parks Method (Autoregressive Model)</td>
<td>1423</td>
</tr>
<tr>
<td>Da Silva Method (Variance-Component Moving Average Model)</td>
<td>1425</td>
</tr>
<tr>
<td>Dynamic Panel Estimator</td>
<td>1427</td>
</tr>
<tr>
<td>Linear Hypothesis Testing</td>
<td>1438</td>
</tr>
</tbody>
</table>
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 in different time periods (years).

The panel data models can be grouped into several categories depending on the 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 and random-effects models
- autoregressive models
- moving average 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 dimension of difference 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 ordinary least squares (OLS) regression is performed.

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 lagged dependent variable.

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

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

- New estimation methods include between estimators, pooled estimators, and dynamic panel estimators that use the generalized method of moments (GMM). 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 creates classification variables that are used in the analysis.

- The TEST statement includes new options for Wald, LaGrange multiplier, and the likelihood ratio tests.

- The new RESTRICT statement specifies linear restrictions on the parameters.

- The FLATDATA statement enables the data to be in a compressed form.

- Several methods that produce heteroscedasticity-consistent (HCCME) and heteroscedasticity- and Autocorrelation-Consistent (HAC) covariance matrices are added because the presence of heteroscedasticity and autocorrelation can result in inefficient and biased estimates of the covariance matrix in the OLS framework.

- Tests are added for poolability, panel stationarity, the existence of cross sectional and time effects, autocorrelation, and cross sectional dependence.

- The LAG statement can generate a large number of missing values, depending on lag order. Typically, it is difficult to create lagged variables in the panel setting. If lagged variables are created in a DATA step, several programming steps that include loops are often needed. By including the LAG statement, the PANEL procedure makes the creation of lagged values easy. The missing values can be replaced with zeros, overall mean, time mean, or cross section mean by using the LAG, ZLAG, XLAG, SLAG, and CLAG statements.

- The OUTPUT statement enables you to output data and estimates that can be used in other analyses.
Getting Started: PANEL Procedure

The following statements use the cost function data from Greene (1990) to estimate the 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. See Greene (1990) for details.

```plaintext
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 1965 7.40245 2.09519 2 1970 7.82644 2.39480
3 1955 8.07153 2.94628 3 1960 8.47679 3.25967
... more lines ...
```

You decide to fit the following model to the data:

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

where \(C_{it}\) and \(P_{it}\) represent the cost and production, and \(v_i\), \(e_t\) and \(\epsilon_{it}\) 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.

```plaintext
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 20.1. A model description is printed first, which reports the estimation method used and the number of cross sections and time periods. The variance components estimates are printed next. Finally, the table of regression parameter estimates shows the estimates, standard errors, and \(t\) tests.
**Figure 20.1** The Variance Components Estimates

**The PANEL Procedure**

**Fuller and Battese Variance Components (RanTwo)**

**Dependent Variable: cost**

<table>
<thead>
<tr>
<th>Model Description</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
<td>RanTwo</td>
</tr>
<tr>
<td>Number of Cross Sections</td>
<td>6</td>
</tr>
<tr>
<td>Time Series Length</td>
<td>4</td>
</tr>
</tbody>
</table>

**Fit Statistics**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>0.3481</td>
</tr>
<tr>
<td>DFE</td>
<td>22</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0158</td>
</tr>
<tr>
<td>Root MSE</td>
<td>0.1258</td>
</tr>
<tr>
<td>R-Square</td>
<td>0.8136</td>
</tr>
</tbody>
</table>

**Variance Component Estimates**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance Component for Cross Sections</td>
<td>0.046907</td>
</tr>
<tr>
<td>Variance Component for Time Series</td>
<td>0.00906</td>
</tr>
<tr>
<td>Variance Component for Error</td>
<td>0.008749</td>
</tr>
</tbody>
</table>

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Hausman Test for Random Effects</td>
<td></td>
</tr>
<tr>
<td>DF</td>
<td>m Value</td>
</tr>
<tr>
<td>----</td>
<td>---------</td>
</tr>
<tr>
<td>1</td>
<td>26.46</td>
</tr>
</tbody>
</table>

| Variable     | DF | Estimate | 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 options;
  FLATDATA options;
  ID cross-section-id time-series-id;
  INSTRUMENTS options;
  LAG options;
  MODEL dependent = regressors < / 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>
<thead>
<tr>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Set Options</td>
<td>PANEL</td>
<td>CORROUT</td>
</tr>
<tr>
<td>Includes correlations in the OUTEST= data set</td>
<td>PANEL</td>
<td>COVOUT</td>
</tr>
<tr>
<td>Includes covariances in the OUTEST= data set</td>
<td>PANEL</td>
<td>DATA=</td>
</tr>
<tr>
<td>Specifies the input data set</td>
<td>FLATDATA</td>
<td>KEEP=</td>
</tr>
<tr>
<td>Specifies variables to keep but not transform</td>
<td>CLASS</td>
<td>OUT=</td>
</tr>
<tr>
<td>Specifies the output data set for CLASS</td>
<td>FLATDATA</td>
<td>OUT=</td>
</tr>
<tr>
<td>STATEMENT</td>
<td>OUTPUT</td>
<td>OUT=</td>
</tr>
<tr>
<td>Specifies the output data set</td>
<td>PANEL</td>
<td>OUTTEST=</td>
</tr>
<tr>
<td>Specifies the name of an output SAS data set</td>
<td>PANEL</td>
<td>OUTTRANS=</td>
</tr>
<tr>
<td>Writes parameter estimates to an output data set</td>
<td>PANEL</td>
<td>PLOTS=</td>
</tr>
<tr>
<td>Writes the transformed series to an output data set</td>
<td>PANEL</td>
<td></td>
</tr>
<tr>
<td>Requests that the procedure produce graphics via the Output Delivery System</td>
<td>PANEL</td>
<td></td>
</tr>
</tbody>
</table>

Declaring the Role of Variables

- Specifies BY-group processing               | BY    
- Specifies the classification variables      | CLASS |
- Transfers the data into uncompressed form    | FLATDATA |
- Specifies the cross section and time ID variables | ID |
- Declares instrumental variables              | INSTRUMENTS |
<table>
<thead>
<tr>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lag Generation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specifies output data set for lags where missing values are replaced with the cross section mean</td>
<td>CLAG</td>
<td>OUT=</td>
</tr>
<tr>
<td>Specifies output data set for lags with missing values included</td>
<td>LAG</td>
<td>OUT=</td>
</tr>
<tr>
<td>Specifies output data set for lags where missing values are replaced with the time period mean</td>
<td>SLAG</td>
<td>OUT=</td>
</tr>
<tr>
<td>Specifies output data set for lags where missing values are replaced with overall mean</td>
<td>XLAG</td>
<td>OUT=</td>
</tr>
<tr>
<td>Specifies output data set for lags where missing values are replaced with zero</td>
<td>ZLAG</td>
<td>OUT=</td>
</tr>
<tr>
<td><strong>Printing Control Options</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prints correlations of the estimates</td>
<td>MODEL</td>
<td>CORRB</td>
</tr>
<tr>
<td>Prints covariances of the estimates</td>
<td>MODEL</td>
<td>COVB</td>
</tr>
<tr>
<td>Suppresses printed output</td>
<td>MODEL</td>
<td>NOPRINT</td>
</tr>
<tr>
<td>Requests that the procedure produce graphics via the Output Delivery System</td>
<td>MODEL</td>
<td>PLOTS=</td>
</tr>
<tr>
<td>Prints fixed effects</td>
<td>MODEL</td>
<td>PRINTFIXED</td>
</tr>
<tr>
<td>Performs tests of linear hypotheses</td>
<td>TEST</td>
<td></td>
</tr>
<tr>
<td><strong>Model Estimation Options</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Requests the $R_p$ statistic for serial correlation under fixed effects</td>
<td>MODEL</td>
<td>BFN</td>
</tr>
<tr>
<td>Requests the Baltagi and Li joint Lagrange multiplier (LM) test for serial correlation and random cross-sectional effects</td>
<td>MODEL</td>
<td>BL91</td>
</tr>
<tr>
<td>Requests the Baltagi and Li LM test for first-order correlation under fixed effects</td>
<td>MODEL</td>
<td>BL95</td>
</tr>
<tr>
<td>Requests the Breusch-Pagan test for one-way random effects</td>
<td>MODEL</td>
<td>BP</td>
</tr>
<tr>
<td>Requests the Breusch-Pagan test for two-way random effects</td>
<td>MODEL</td>
<td>BP2</td>
</tr>
<tr>
<td>Requests the Bera, Sosa Escudero, and Yoon modified Rao’s score test</td>
<td>MODEL</td>
<td>BSY</td>
</tr>
<tr>
<td>Specifies the between-groups model</td>
<td>MODEL</td>
<td>BTWNG</td>
</tr>
<tr>
<td>Specifies the between-time-periods model</td>
<td>MODEL</td>
<td>BTWNT</td>
</tr>
<tr>
<td>Requests the Berenblut-Webb statistic for serial correlation under fixed effects</td>
<td>MODEL</td>
<td>BW</td>
</tr>
<tr>
<td>Requests cross-sectional dependence tests</td>
<td>MODEL</td>
<td>CDTEST</td>
</tr>
<tr>
<td>Requests the clustered HCCME estimator for the covariance matrix</td>
<td>MODEL</td>
<td>CLUSTER</td>
</tr>
<tr>
<td>Specifies the Da Silva method</td>
<td>MODEL</td>
<td>DASILVA</td>
</tr>
<tr>
<td>Requests the Durbin-Watson statistic for serial correlation under fixed effects</td>
<td>MODEL</td>
<td>DW</td>
</tr>
<tr>
<td>Specifies the one-way fixed-effects model</td>
<td>MODEL</td>
<td>FIXONE</td>
</tr>
<tr>
<td>Description</td>
<td>Statement</td>
<td>Option</td>
</tr>
<tr>
<td>-----------------------------------------------------------------------------</td>
<td>-----------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Specifies the one-way fixed-effects model with respect to time</td>
<td>MODEL</td>
<td>FIXONETIME</td>
</tr>
<tr>
<td>Specifies the two-way fixed-effects model</td>
<td>MODEL</td>
<td>FIXTWO</td>
</tr>
<tr>
<td>Specifies the first-differenced methods for one-way models</td>
<td>MODEL</td>
<td>FDONE</td>
</tr>
<tr>
<td>Specifies the first-differenced methods for one-way models with respect to time</td>
<td>MODEL</td>
<td>FDONETIME</td>
</tr>
<tr>
<td>Specifies the first-differenced methods for two-way models</td>
<td>MODEL</td>
<td>FDTWO</td>
</tr>
<tr>
<td>Specifies the Moore-Penrose generalized inverse</td>
<td>MODEL</td>
<td>GINV = G4</td>
</tr>
<tr>
<td>Requests the Gourieroux, Holly, and Monfort test for two-way random effects</td>
<td>MODEL</td>
<td>GHM</td>
</tr>
<tr>
<td>Specifies the dynamic panel estimator model (one-step GMM)</td>
<td>MODEL</td>
<td>GMM1</td>
</tr>
<tr>
<td>Specifies the dynamic panel estimator model (two-step GMM)</td>
<td>MODEL</td>
<td>GMM2</td>
</tr>
<tr>
<td>Requests the HAC estimator for the variance-covariance matrix</td>
<td>MODEL</td>
<td>HAC=</td>
</tr>
<tr>
<td>Requests the HCCME estimator for the covariance matrix</td>
<td>MODEL</td>
<td>HCCME=</td>
</tr>
<tr>
<td>Requests the Honda test for one-way random effects</td>
<td>MODEL</td>
<td>HONDA</td>
</tr>
<tr>
<td>Requests the Honda test for two-way random effects</td>
<td>MODEL</td>
<td>HONDA2</td>
</tr>
<tr>
<td>Specifies the dynamic panel estimator model (iterated GMM)</td>
<td>MODEL</td>
<td>ITGMM</td>
</tr>
<tr>
<td>Requests the King and Wu test for two-way random effects</td>
<td>MODEL</td>
<td>KW</td>
</tr>
<tr>
<td>Specifies the order of the moving average error process for Da Silva method</td>
<td>MODEL</td>
<td>M=</td>
</tr>
<tr>
<td>Suppresses the intercept term</td>
<td>MODEL</td>
<td>NOINT</td>
</tr>
<tr>
<td>Specifies the Parks method</td>
<td>MODEL</td>
<td>PARKS</td>
</tr>
<tr>
<td>Prints the Φ matrix for Parks method</td>
<td>MODEL</td>
<td>PHI</td>
</tr>
<tr>
<td>Specifies the pooled model</td>
<td>MODEL</td>
<td>POOLED</td>
</tr>
<tr>
<td>Requests poolability tests for one-way fixed effects and pooled model</td>
<td>MODEL</td>
<td>POOLTEST</td>
</tr>
<tr>
<td>Specifies the one-way random-effects model</td>
<td>MODEL</td>
<td>RANONE</td>
</tr>
<tr>
<td>Specifies the two-way random-effects model</td>
<td>MODEL</td>
<td>RANTWO</td>
</tr>
<tr>
<td>Prints autocorrelation coefficients for Parks method</td>
<td>MODEL</td>
<td>RHO</td>
</tr>
<tr>
<td>Controls the check for singularity</td>
<td>MODEL</td>
<td>SINGULAR=</td>
</tr>
<tr>
<td>Specifies the method for panel unit root/stationarity test</td>
<td>MODEL</td>
<td>UROOTTEST=</td>
</tr>
</tbody>
</table>
## 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 1467 for details about the structure of the OUTEST= data set.

### OUTTRANS= SAS-data-set

names an output data set to contain the transformed series for further analysis and computation of models with time observations greater than two. See the section “OUTTRANS= Data Set” on page 1468 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 1467 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 1467 for details.

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

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.

PREDSENSURFACE produces a surface plot of predicted values.

QQ produces a QQ 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 1465.

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, PARKS, DASILVA, NOINT, NOPRINT, PRINTFIXED, M=, PHI, RHO, VCOMP=, and SINGULAR=. When specified in the PROC PANEL statement, these options are equivalent to specifying the options for every MODEL statement. See the section “MODEL Statement” on page 1389 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:

```sas
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.

In PROC PANEL, the CLASS statement enables you to output class variables to a data set that contains a copy of the original data.

### FLATDATA Statement

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

The following options must be specified in the FLATDATA statement:

- **BASE=(variable, variable, . . . , variable)** 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 basename, 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 statements:

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

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

---

**INSTRUMENTS Statement**

INSTRUMENTS options ;

The INSTRUMENTS statement denotes which variables are used in the moment condition equations of the dynamic panel estimator. 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)
DIFFERENCEDEQ=(variable, 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)
LEVELSEQ=(variable, 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.

Because a variable can be used as an instrument only if it is either exogenous or predetermined, 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 to 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;
  inst 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, see “Dynamic Panel Estimator” on page 1427.

Note that for each MODEL statement, one INSTRUMENT statement is required. In other words, if there are two models to be estimated by using GMM1 within one PANEL procedure, then there should be two INSTRUMENT statements. For example,

```
proc panel data=test;
  inst depvar pred=(x1 x2) exog=(x3 x4 x5) correlated=(x3 x4 x5);
  model y = y_1 x1 x2 / gmm1 maxband=6 nolevels ginv=g4 artest=5;
  inst pred=(x2 x4) exog=(x3 x5) correlated=(x3 x4);
  model y = y_1 x2 / gmm1 maxband=6 nolevels ginv=g4 artest=5;
  id cs ts;
run;
```
LAG, ZLAG, XLAG, SLAG, or CLAG Statement

**LAG**  \( \text{var}_{1} (\text{lag}_{1} \text{ lag}_{2} \ldots \text{lag}_{T}) , \ldots , \text{var}_{N} (\text{lag}_{1} \text{ lag}_{2} \ldots \text{lag}_{T}) < / \text{OUT} = \text{SAS-data-set} > ; \)

Generally, creating lags of variables in a panel setting is a tedious process in which you must generate many DATA step statements. The PANEL procedure now enables you to generate lags of any series without jumping across the boundary of any individual series. The LAG statement is a data set generation tool. Using the data created by a LAG statement requires a subsequent PROC PANEL call. You can specify more than one LAG statement in each call to PROC PANEL.

You must specify the OUT= option in the LAG statement. The output data set includes all variables in the input set, plus the lags that are denoted with the convention \( \text{var}_{\text{lag}} \). 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 instead of LAG with otherwise identical syntax:

**CLAG**  \( \text{var}_{1} (\text{lag}_{1} \text{ lag}_{2} \ldots \text{lag}_{T}) , \ldots , \text{var}_{N} (\text{lag}_{1} \text{ lag}_{2} \ldots \text{lag}_{T}) < / \text{OUT} = \text{SAS-data-set} > ; \)

replaces missing values with the cross section mean for that variable in that cross section. Missing values are replaced only if they are in the generated (lagged) series. Missing variables in the original variables are not changed.

**SLAG**  \( \text{var}_{1} (\text{lag}_{1} \text{ lag}_{2} \ldots \text{lag}_{T}) , \ldots , \text{var}_{N} (\text{lag}_{1} \text{ lag}_{2} \ldots \text{lag}_{T}) < / \text{OUT} = \text{SAS-data-set} > ; \)

replaces missing values with the time mean for that variable in that time period. Missing values are replaced only if they are in the generated (lagged) series. Missing variables in the original variables are not changed.

**XLAG**  \( \text{var}_{1} (\text{lag}_{1} \text{ lag}_{2} \ldots \text{lag}_{T}) , \ldots , \text{var}_{N} (\text{lag}_{1} \text{ lag}_{2} \ldots \text{lag}_{T}) < / \text{OUT} = \text{SAS-data-set} > ; \)

replaces missing values with the overall mean for that variable. Missing values are replaced only if they are in the generated (lagged) series. Missing variables in the original variables are not changed.

**ZLAG**  \( \text{var}_{1} (\text{lag}_{1} \text{ lag}_{2} \ldots \text{lag}_{T}) , \ldots , \text{var}_{N} (\text{lag}_{1} \text{ lag}_{2} \ldots \text{lag}_{T}) < / \text{OUT} = \text{SAS-data-set} > ; \)

replaces missing values with 0 for that variable. Missing values are replaced only if they are in the generated (lagged) series. Missing variables in the original variables are not changed.

Assume that data set A has been sorted by cross section and by time period within cross section (or that the FLATDATA statement has been specified) and that the variables are \( Y, X_1, X_2, \) and \( X_3 \). The following PROC PANEL statements generate a series with lags 1 and 3 of the \( X_1 \) variable; lags 3, 6, and 9 of the \( X_2 \) variable; and lag 2 of the \( X_3 \) 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 a zeroing instead of missing values, then you specify the following:

```
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**

```plaintext
MODEL response = regressors / options;
```

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

The error structure is specified by the PARKS, DASILVA, FIXONE, FIXONETIME, FIXTWO, FDONE, FDONETIME, FDTWO, RANONE, RANTWO, GMM1, GMM2, and ITGMM options. More than one of these options can be used, in which case the analysis is repeated for each error structure model specified.

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

```
label : MODEL ...;
```

The following `options` can be specified in the MODEL statement after a slash `/`.

- **ARTEST=integer**
  specifies the maximum order of the test for the presence of AR effects in the residual in the dynamic panel model. The acceptable range of values for this option is 1 to \( t - 3 \).

- **ATOL=number**
  specifies the convergence criterion for iterated GMM when convergence of the method is determined by convergence in the weighting matrix. The convergence criterion must be positive. The default option is the BTOL= option unless the ATOL= option is specified. See the section “Dynamic Panel Estimator” on page 1427 for details.

- **BANDOPT=TRAILING | CENTERED | LEADING**
  specifies which observations are included in the instrument list when the MAXBAND= option is specified. This option should be used only for exogenous instruments. BANDOPT=TRAILING is the default. See the section “Dynamic Panel Estimator” on page 1427 for details.

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

- **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. For more information, see the section “Dynamic Panel Estimator” on page 1427.

- **BL91**
  requests the Baltagi and Li (1991) joint 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.

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 must be positive. The default is BTOL=1E–8. See the section “Dynamic Panel Estimator” on page 1427 for details.

BTWNG specifies that a between-groups model be estimated.

BTWNT specifies that a between-time-periods model be estimated.

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 zero.

CLUSTER specifies the cluster correction for the covariance matrix. The cluster correction can be requested with HCCME=0, 1, 2 or 3.

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

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

DASILVA specifies that the model be estimated by using the Da Silva method, which assumes a mixed variance-component moving average model for the error structure. See the section “Da Silva Method (Variance-Component Moving Average Model)” on page 1425 for details.
**Model Statement**

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

**FDONE**  
requests that a one-way model be estimated by using first-differenced methods.

**FDONETIME**  
requests that a one-way model that corresponds to time effects be estimated by using first-differenced methods.

**FDTWO**  
requests that a two-way model be estimated by using first-differenced methods.

**FIXONE**  
specifies that a one-way fixed-effects model be estimated with the one-way model corresponding to cross-sectional effects only.

**FIXONETIME**  
specifies that a one-way fixed-effects model be estimated with the one-way model corresponding to time effects only.

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

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

**GINV=** *G2 | G4*  
specifies what type of generalized inverse to use. The default is a G2 inverse. The G4 inverse is generally more desirable except that it is a more numerically intensive methodology.

**GMM1**  
requests that the model be estimated in a single step by using the dynamic panel estimator method, which allows for autoregressive processes. When you specify this option, you must specify one **INSTRUMENT** statement for each **MODEL** statement. For more information, see the section “Dynamic Panel Estimator” on page 1427.

**GMM2**  
requests that the model be estimated in two steps by using the dynamic panel estimator method. An initial first step is used to form an estimator for the weighting matrix that is used in the second step. For more information, see the section “Dynamic Panel Estimator” on page 1427.

**HAC <(hac-options)>**  
specifies the heteroscedasticity- and autocorrelation-consistent (HAC) covariance matrix estimator. This option is not available for between models and cannot be specified with the HCCME option. When you specify this option, you can also specify the following **hac-options** within parentheses:

**KERNEL=value**  
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 Turkey-Hanning kernel function.
TRUNCATED specifies the truncated kernel function.

The default is KERNEL=TRUNCATED.

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

BANDWIDTH=value
specifies the fixed bandwidth value or bandwidth selection method which is used in the kernel function. You can specify the following values:

ANDREWS91 | ANDREWS

NEWEYWEST94<(C=number)>
NW94 <(C=number)>
specifies the Newey and West (1994) bandwidth selection method. The C= option can be specified within parentheses for the calculation of lag selection parameter; the default is C=12.

SAMPLESIZE<(option-list)>
SS<(option-list)>
specifies that the bandwidth be calculated 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 \( \gamma, r \) and \( c \) are values specified by the following options within parentheses and separated by commas.

GAMMA=number
specifies the coefficient \( \gamma \) in the equation. The default is \( \gamma = 0.75 \).

RATE=number
specifies the growth rate \( r \) in the equation. The default is \( r = 0.3333 \).

CONSTANT=number
specifies the constant \( c \) in the equation. The default is \( c = 0.5 \).

INT
specifies that the bandwidth parameter must be integer; that is, \( b = [\gamma T^r + c] \), where \( [x] \) denotes the largest integer less than or equal to \( x \).
number

specifies the fixed value of the bandwidth parameter.

The default is BANDWIDTH=ANDREWS91.

PREWHITENING

specifies that prewhitening is required in the covariance calculation.

ADJUSTDF

specifies that the adjustment of degrees of freedom is required in the covariance calculation.

See the section “Heteroscedasticity- and Autocorrelation-Consistent Covariance Matrices” on page 1442 for details.

HCCME= NO | number

specifies the type of HCCME covariance matrix requested. If you specify HCCME=NO, the covariance matrix is not corrected. The value number can be any integer from 0 to 4, inclusive. See the section “Heteroscedasticity-Corrected Covariance Matrices” on page 1439 for details. By default, HCCME=NO.

HONDA

requests the Honda one-way test for random effects.

HONDA2

requests the Honda two-way test for random effects.

ITGMM

specifies that the model be estimated by using the dynamic panel estimator method, but that PROC PANEL keep updating the weighting matrix until either the parameter vector converges or the weighting matrix converges. See the section “Dynamic Panel Estimator” on page 1427 for details.

ITPRINT

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

KW

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

M=number

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

MAXBAND=integer

specifies the maximum number of time periods (per instrumental variable) that are allowed into the moment condition. The acceptable range of values for this option is 1 to $T - 1$. 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. See the section “Dynamic Panel Estimator” on page 1427 for details.

MAXITER=integer

specifies the maximum number of iterations allowed for the iterated GMM option. The default value is MAXITER=200. See the section “Dynamic Panel Estimator” on page 1427 for details.
specifies the well-known Newey-West estimator, a special HAC estimator with (1) the Bartlett kernel, (2) the bandwidth parameter 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 \( 0.75T^{0.3333} + 0.5 \), as shown in the equation (15.17) in Stock and Watson (2002). When you specify COVEST=NEWEYWEST, you can specify the following options in parentheses and separate them with commas:

\[ \text{GAMMA=} \text{number} \]

specifies the coefficient \( \gamma \) in the equation. The default is \( \gamma = 0.75 \).

\[ \text{RATE=} \text{number} \]

specifies the growth rate \( r \) in the equation. The default is \( r = 0.3333 \).

\[ \text{CONSTANT=} \text{number} \]

specifies the constant \( c \) in the equation. The default is \( c = 0.5 \).

NODIFFS

specifies that the dynamic panel model be estimated without moment conditions from the difference equations. See the section “Dynamic Panel Estimator” on page 1427 for details.

NOESTIM

limits the estimation of a FIXONE, FIXONETIME, 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.

NOLEVELS

specifies that the dynamic panel model be estimated without moment conditions from the level equations. See the section “Dynamic Panel Estimator” on page 1427 for details.

NOPRINT

suppresses the normal printed output.

PARKS

specifies that the model be estimated by using the Parks method, which assumes a first-order autoregressive model for the error structure. See the section “Parks Method (Autoregressive Model)” on page 1423 for details.

PHI

prints the \( \Phi \) matrix of estimated covariances of the observations for the Parks method. The PHI option is relevant only when the PARKS option is used. See the section “Parks Method (Autoregressive Model)” on page 1423 for details.

POOLED

specifies that a pooled (OLS) model be estimated.

POOLTEST

requests poolability tests for one-way fixed effects and pooled models.
PRINTFIXED
prints the fixed effects.

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

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

RHO
prints the estimated autocorrelation coefficients for the Parks method.

ROBUST
specifies that the robust weighting matrix be used in the calculation of the covariance matrix of the single-step, two-step and iterated GMM dynamic panel estimator. See the section “Dynamic Panel Estimator” on page 1427 for details.

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

TIME
specifies that the model be estimated by using the dynamic panel estimator method, but that PROC PANEL includes time dummy variables to model any time effects present in the data. See the section “Dynamic Panel Estimator” on page 1427 for details.

UROOTTEST((test1< (test-options) , test2< (test-options)> . . . < option1 < option2. . . > )
STATIONARITY((test1< (test-options) , test2< (test-options)> . . . < option1 < option2. . . > )
specifies tests of stationarity or unit root for panel data and the options for each test. The panel unit root test (or stationarity test) will test the existence of unit root for the dependent variables only. Six tests are available. You can specify all or some of these tests, separated by commas. 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.

ALL
requests that all panel unit root and stationarity tests be performed.

BREITUNG< (test-options) >
requests Breitung’s unbiased test, t test and 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). The following test-options are available for this test:

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

LAGS=type | value
specifies the method to choose the lag order for the augmented Dickey-Fuller (ADF) regressions. You can specify a value for the order of lags. If the specified lag order is too big to run linear regression (LAGS > T − k, where T is the number of observations and k is the number of parameters), then the lag order is set to \( \left\lfloor \frac{12(T/100)^{1/4}}{1} \right\rfloor \) or T − k − 1, whichever is smaller. Alternatively, you can specify one of the following types:
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 AIC.

SBC

SIC

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

The default is LAGS=MAIC.

MAXLAGS=value
specifies the maximum lag order that the model allows. The default value is \[12(T/100)^{1/4}\].
If value is larger than 0 and larger than \(T - k\), then the maximum lag order is set to the
default value of \[12(T/100)^{1/4}\] or \(T - k - 1\), whichever is smaller. This option is ignored
if you specify LAGS=value.

COMBINATION < (test-options) >
FISHER < (test-options) >
specifies combination tests 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:

TEST=ADF | PP
selects the time series unit root test for combination tests (Fisher’s test). ADF specifies the
augmented Dickey-Fuller (ADF) test, and ignores the BANDWIDTH and KERNEL options
for the combination tests. PP specifies the Phillips and Perron (1988) unit root test. When
you specifies TEST = PP, the LAGS and MAXLAGS options are ignored for the combination
tests.
The default is TEST=PP.

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

BARTLETT specifies the Bartlett kernel function.
PARZEN specifies the Parzen kernel function.
MODEL Statement

QS specifies the quadratic spectral kernel function.
TH specifies the Turkey-Hanning kernel function.
TRUNCATED specifies the truncated kernel function.

The default is KERNEL=QS.

BANDWIDTH=ANDREWS | number

specifies the bandwidth for the kernel. If you specify BANDWIDTH=ANDREWS, the bandwidth is selected by the Andrews method. If you specify a nonnegative number, the bandwidth is set to that value. The default is BANDWIDTH=ANDREWS.

DETAIL

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

LAGS=type | value

specifies the method to choose the lag order for the augmented Dickey-Fuller (ADF) regressions. You can specify a value for the order of lags. If the specified lag order is too big to run linear regression (LAGS > T – k, where T is the number of observations 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. Alternatively, you can specify one of the following types:

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 AIC.

SBC

SIC

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

The default is LAGS=MAIC.

MAXLAGS=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 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 LAGS=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=value
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 Turkey-Hanning kernel function.
TRUNCATED specifies the truncated kernel function.

The default is KERNEL=QS.

BANDWIDTH=ANDREWS | number
specifies the bandwidth for the kernel. If you specify BANDWIDTH=ANDREWS, the bandwidth is selected with the Andrews method. If you specify a nonnegative number, the bandwidth is set to that value. The default is BANDWIDTH=ANDREWS.

HT
specifies 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 the following test-options:

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

LAGS=type | value
specifies the method to choose the lag order for the augmented Dickey-Fuller (ADF) regressions. You can specify a value for the order of lags. If the specified lag order is too big to run linear regression (LAGS > T – k, where T is the number of observations and k is the number of parameters), then the lag order is set to \( 12(T/100)^{1/4} \) or \( T – k – 1 \), whichever is smaller. Alternatively, you can specify one of the following types:

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 AIC.

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

The default is LAGS=MAIC.

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

LLC < (test-options) >
specifies the Levin, Lin, and Chu (2002) panel unit root 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=value
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 Turkey-Hanning kernel function.
TRUNCATED specifies the truncated kernel function.

The default is KERNEL=QS.

BANDWIDTH=ANDREWS | number
specifies the bandwidth for the kernel. If you specify BANDWIDTH=ANDREWS, the bandwidth is selected with the Andrews method. If you specify a nonnegative number, the bandwidth is set to that value. The default is BANDWIDTH=LLCBAND, where the bandwidth is set to be \( \tilde{k} = \left[ 3.21T^{\frac{1}{3}} \right] \), according to Levin, Lin, and Chu (2002).
LAGS=type | value
specifies the method to choose the lag order for the augmented Dickey-Fuller (ADF) regressions. You can specify a value for the order of lags. If the specified lag order is too big to run linear regression (LAGS > T – k, where T is the number of observations and k is the number of parameters), then the lag order is set to 12(T/100)^{1/4} or T – k – 1, whichever is smaller. Alternatively, you can specify one of the following types:

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 AIC.

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

The default is LAGS=MAIC.

MAXLAGS=value
specifies the maximum lag order that the model allows. The default value is 12(T/100)^{1/4}. If value is larger than 0 and larger than T – k, then the maximum lag order is set to be the default value of 12(T/100)^{1/4} or T – k – 1, whichever is smaller. This option is ignored if you specify LAGS=value.

Two tests, LLC and BREITUNG’s, are specified in the following UROOTTEST option specification:
urootest = (llc=(kernel=parzen lags=aic), breitung=(lags=gs ) maxlags=2 kernel=bartlett)

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’s test, the lag order is GS with a maximum lag order 2. The KERNEL option is ignored by BREITUNG’s test because it is not a valid option.

VCOMP=FB | NL | WH | WK
specifies the type of variance component estimate to use. The default is VCOMP=FB for balanced data and VCOMP=WK for unbalanced data. See the section “One-Way Random-Effects Model” on page 1415 and “Two-Way Random-Effects Model” on page 1418 for details.
WOOLDRIDGE02 requests the Wooldridge (2002) test for the presence of unobserved effects.

**OUTPUT Statement**

```
OUTPUT OUT=SAS-data-set = 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.

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. 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.

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**

```plaintext
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. TEST and RESTRICT statements before the first MODEL statement are automatically associated with the first MODEL statement, in addition to any TEST and RESTRICT 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:

```plaintext
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. The 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 presented, PROC PANEL uses the quoted string. In the event no label is present, PROC PANEL automatically labels the tests. If both a TEST and a RESTRICT statement are specified, the test is run with restrictions applied.

Note that for the DaSilva method, 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:

```sas
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 shows the correct syntax:

```sas
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 is 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 (say `state`). The time observations for the Y variable are recorded horizontally. So the variable Y_1 is the first period’s time observation, Y_10 is the tenth period’s observation for some state. The same holds for the other variables. You have 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:

```sas
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 1385 and Example 20.2 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:

```sas
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:

```sas
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 20.1 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. 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 \( p + \frac{p(p + 1)}{2} \) 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**

A consequence of estimating a linear model with a restriction is that the error degrees of freedom increase by the number of restrictions. PROC PANEL produces the Lagrange multiplier associated with each restriction.

Say 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:

\[
R_{1,1}\beta_1 + R_{1,2}\beta_2 + \cdots + R_{1,p}\beta_p = q_1
\]

\[
R_{2,1}\beta_1 + R_{2,2}\beta_2 + \cdots + R_{2,p}\beta_p = q_2
\]

\[\vdots\]

\[
R_{r,1}\beta_1 + R_{r,2}\beta_2 + \cdots + R_{r,p}\beta_p = q_r
\]
To economize on notation, you can represent the restriction structure in the following matrix notation $R\beta = q$. The restricted $\beta$ estimator is given by:

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

The LaGrange multipliers are given as:

$$\lambda_* = \left[ R(X'X)^{-1}R' \right]^{-1} (R\beta - q)$$

The standard errors of the LaGrange Multipliers are calculated from the following relationship:

$$\text{Var}(\lambda_*) = \left[ R(X'X)^{-1}R' \right]^{-1} R \text{Var}(\beta) R' \left[ R(X'X)^{-1}R' \right]^{-1}$$

A significant LaGrange multiplier implies that you can reject the null hypothesis that the restrictions are 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.

**Notation**

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

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

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 $X$ 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 \( \gamma_i \)s are nonrandom parameters to be estimated.

Let \( Q_0 = \text{diag}(E_{T_i}) \), with \( \bar{J}_{T_i} = J_{T_i}/T_i \) and \( E_{T_i} = 1_{T_i} - \bar{J}_{T_i} \), where \( J_{T_i} \) is a matrix of \( T_i \) ones.

The matrix \( Q_0 \) 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 \( \tilde{x}_{it} \) is a row of the general matrix \( X_s \), where the subscripted \( s \) implies the constant (column of ones) is missing.

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

\[ \tilde{\beta}_s = (\tilde{X}_s' \tilde{X}_s)^{-1} \tilde{X}_s' \tilde{y} \]

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:

\[ \gamma_i = \tilde{y}_i - \tilde{\beta}_s \tilde{x}_i, \quad \text{for} \quad i = 1 \ldots N \]

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 \( \gamma_N \).

The within model sum of squared errors is:

\[ \text{SSE} = \sum_{i=1}^{N} \sum_{t=1}^{T_i} (y_{it} - \gamma_i - X_s \tilde{\beta}_s)^2 \]

The estimated error variance can be written:

\[ \hat{\sigma}_e^2 = \text{SSE}/(M - N - (K - 1)) \]

Alternatively, an equivalent way to express the error variance is

\[ \hat{\sigma}_e^2 = \tilde{u}' Q_0 \tilde{u}/(M - N - (K - 1)) \]

where the residuals \( \tilde{u} \) are given by \( \tilde{u} = (I_M - j_M j_M'/M)(y - X_s \tilde{\beta}_s) \) if there is an intercept and by \( \tilde{u} = (y - X_s \tilde{\beta}_s) \) 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:

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

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:**
  
  $$\text{Var} [y_i] = \text{Var} [D_i] = \frac{\hat{\sigma}_e^2}{T_i} + \tilde{x}_i' \text{Var} \left[ \tilde{\beta}_s \right] \tilde{x}_i.$$  
  $$\text{Cov} [y_i, y_j] = \text{Cov} [D_i, D_j] = \tilde{x}_i' \text{Var} \left[ \tilde{\beta}_s \right] \tilde{x}_j.$$  
  $$\text{Cov} [y_i, \tilde{\beta}_s] = \text{Cov} [D_i, \tilde{\beta}_s] = -\tilde{x}_i' \text{Var} \left[ \tilde{\beta}_s \right]$$

- **intercept:**
  
  $$\text{Var} [D_i] = \frac{\hat{\sigma}_e^2}{T_i} + \frac{\hat{\sigma}_e^2}{T_N} + (\tilde{x}_i - \tilde{x}_N)' \text{Var} \left[ \tilde{\beta}_s \right] (\tilde{x}_i - \tilde{x}_N).$$  
  $$\text{Cov} [D_i, D_j] = \frac{\hat{\sigma}_e^2}{T_N} + (\tilde{x}_i - \tilde{x}_N)' \text{Var} \left[ \tilde{\beta}_s \right] (\tilde{x}_j - \tilde{x}_N).$$  
  $$\text{Var} [\text{Intercept}] = \text{Var} [y_N] = \frac{\hat{\sigma}_e^2}{T_N} + \tilde{x}_N' \text{Var} \left[ \tilde{\beta}_s \right] \tilde{x}_N.$$  
  $$\text{Cov} [D_i, \tilde{\beta}_s] = -(\tilde{x}_i - \tilde{x}_N)' \text{Var} \left[ \tilde{\beta}_s \right]$$  
  $$\text{Cov} [\text{Intercept}, D_i] = -\frac{\hat{\sigma}_e^2}{T_i} + \tilde{x}_i' \text{Var} \left[ \tilde{\beta}_s \right] (\tilde{x}_i - \tilde{x}_N).$$  
  $$\text{Cov} [\text{Intercept}, \tilde{\beta}_s] = -\tilde{x}_N' \text{Var} \left[ \tilde{\beta}_s \right]$$

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

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

where the $y_i$s and $\alpha_t$s 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

Assume that the data are balanced (for example, all cross sections have $T$ observations). Then you can write the following:

$$
\tilde{y}_{it} = y_{it} - \bar{y}_i - \bar{y}_t + \bar{y}
$$

$$
\tilde{x}_{it} = x_{it} - \bar{x}_i - \bar{x}_t + \bar{x}
$$

where the symbols:

- $y_{it}$ and $x_{it}$ are the dependent variable (a scalar) and the explanatory variables (a vector whose columns are the explanatory variables not including a constant), respectively
- $\bar{y}_i$ and $\bar{x}_i$ are cross section means
- $\bar{y}_t$ and $\bar{x}_t$ are time means
- $\bar{y}$ and $\bar{x}$ are the overall means

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

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

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 $\gamma$ and the time effects by $\alpha$. These effects are calculated from the following relations:

$$
\hat{\gamma}_i = (\tilde{y}_i - \bar{y}) - \tilde{\beta}_x (\tilde{x}_i - \bar{x})
$$

$$
\hat{\alpha}_t = (\tilde{y}_t - \bar{y}) - \tilde{\beta}_x (\tilde{x}_t - \bar{x})
$$

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:

$$
D^C_i = \hat{\gamma}_i + \hat{\alpha}_T
$$

$$
D^T_i = \hat{\alpha}_i - \hat{\alpha}_T
$$
When an intercept is specified, the equations for dummy variables and intercept are:
\[
\begin{align*}
D^C_i &= \hat{y}_t - \hat{y}_N \\
D^T_t &= \hat{\alpha}_t - \hat{\alpha}_T \\
\text{Intercept} &= \hat{y}_N + \hat{\alpha}_T
\end{align*}
\]

The sum of squared errors is:
\[
\text{SSE} = \sum_{i=1}^{N} \sum_{t=1}^{T_i} (y_{it} - \gamma_i - \alpha_t - X_s \hat{\beta}_s)^2
\]

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 \( \hat{\beta}_s \) is given by:
\[
\text{Var} \left[ \hat{\beta}_s \right] = \hat{\sigma}_\epsilon^2 (\bar{X}_s' \bar{X}_s)^{-1}
\]

**Variance Covariance of Dummy Variables with No Intercept**

The variances and covariances of the dummy variables are given with the NOINT specification as follows:
\[
\begin{align*}
\text{Var} \left( D^C_i \right) &= \hat{\sigma}_\epsilon^2 \left( \frac{1}{T} + \frac{1}{N} - \frac{1}{NT} \right) \\
&\quad + (\bar{x}_{i.} + \bar{x}_{.T} - \bar{x})' \text{Var} \left[ \hat{\beta}_s \right] (\bar{x}_{i.} + \bar{x}_{.T} - \bar{x}) \\
\text{Var} \left( D^T_t \right) &= \frac{2\hat{\sigma}_\epsilon^2}{N} + (\bar{x}_t - \bar{x}_{.T})' \text{Var} \left[ \hat{\beta}_s \right] (\bar{x}_t - \bar{x}_{.T}) \\
\text{Cov} \left( D^C_i, D^C_j \right) &= \hat{\sigma}_\epsilon^2 \left( \frac{1}{N} - \frac{1}{NT} \right) \\
&\quad + (\bar{x}_{i.} + \bar{x}_{.t} - \bar{x})' \text{Var} \left[ \hat{\beta}_s \right] (\bar{x}_{j.} + \bar{x}_{.t} - \bar{x}) \\
\text{Cov} \left( D^T_t, D^T_u \right) &= \frac{\hat{\sigma}_\epsilon^2}{N} + (\bar{x}_t - \bar{x}_{.T})' \text{Var} \left[ \hat{\beta}_s \right] (\bar{x}_u - \bar{x}_{.T}) \\
\text{Cov} \left( D^C_i, D^T_t \right) &= -\frac{\hat{\sigma}_\epsilon^2}{N} + (\bar{x}_{i.} + \bar{x}_{.t} - \bar{x})' \text{Var} \left[ \hat{\beta}_s \right] (\bar{x}_t - \bar{x}_{.T}) \\
\text{Cov} \left( D^C_i, \beta \right) &= - (\bar{x}_{i.} + \bar{x}_{.t} - \bar{x})' \text{Var} \left[ \hat{\beta}_s \right] \\
\text{Cov} \left( D^T_t, \beta \right) &= - (\bar{x}_t - \bar{x}_{.T})' \text{Var} \left[ \hat{\beta}_s \right]
\end{align*}
\]
Variance Covariance of Dummy Variables with Intercept

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

$$\text{Var} \left( D_i^C \right) = \frac{2\sigma^2}{T} + \left( \bar{x}_i - \bar{x} \right)' \text{Var} \left[ \hat{\beta}_s \right] (\bar{x}_i - \bar{x})$$

$$\text{Var} \left( D_i^T \right) = \frac{2\sigma^2}{N} + \left( \bar{x}_i - \bar{x}_T \right)' \text{Var} \left[ \hat{\beta}_s \right] (\bar{x}_i - \bar{x}_T)$$

$$\text{Var} \left( \text{Intercept} \right) = \sigma^2 \left( \frac{1}{T} + \frac{1}{N} - \frac{1}{NT} \right) + \left( \bar{x}_N + \bar{x}_T - \bar{x} \right)' \text{Var} \left[ \hat{\beta}_s \right] (\bar{x}_N + \bar{x}_T - \bar{x})$$

$$\text{Cov} \left( D_i^C, D_j^C \right) = \frac{\sigma^2}{T} + \left( \bar{x}_i - \bar{x}_N \right)' \text{Var} \left[ \hat{\beta}_s \right] (\bar{x}_j - \bar{x}_N)$$

$$\text{Cov} \left( D_i^T, D_j^T \right) = \frac{\sigma^2}{N} + \left( \bar{x}_i - \bar{x}_T \right)' \text{Var} \left[ \hat{\beta}_s \right] (\bar{x}_j - \bar{x}_T)$$

$$\text{Cov} \left( D_i^C, \text{Intercept} \right) = \left( \frac{\sigma^2}{T} \right) + \left( \bar{x}_i - \bar{x}_N \right)' \text{Var} \left[ \hat{\beta}_s \right] (\bar{x}_N + \bar{x}_T - \bar{x})$$

Unbalanced Panels

Let $X_*$ and $y_*$ 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 $D_t$ 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

$$Z = (Z_1, Z_2)$$

where $Z_1 = (D_1, D_2, \ldots, D_T)'$ and $Z_2 = \text{diag}(D_{1jN}, D_{2jN}, \ldots, D_{TjN})$. The matrix $Z$ gives the dummy variable structure for the two-way model.

Let

$$\Delta_N = Z_1' Z_1$$

$$\Delta_T = Z_2' Z_2$$

$$A = Z_2' Z_1$$

$$\bar{Z} = Z_2 - Z_1 \Delta_N^{-1} A'$$

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

$$P = (I_M - Z_1 \Delta_N^{-1} Z_1') - \bar{Z} Q^{-1} \bar{Z}'$$
The estimate of the regression slope coefficients is given by

$$\tilde{\beta}_s = (X'_{s*}PX_{s*})^{-1}X'_{s*}Py_*$

where $X_{s*}$ is the $X_*$ matrix without the vector of 1s.

The estimator of the error variance is

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

where the residuals are given by $\tilde{u} = (I_M - \bar{X}'_M/M)(y_* - X_{s*}\tilde{\beta}_s)$ if there is an intercept in the model and by $\tilde{u} = y_* - X_{s*}\tilde{\beta}_s$ if there is no intercept.

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

$$\tilde{v} = Pv$$

The variable being transformed is $v$, which could be $y$ or any column of $X$. 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 $y$ is the column vector of cross-sectional effects and $\alpha$ is the column vector of time effects,

$$\tilde{\alpha} = Q^{-1}\tilde{Z}'y - Q^{-1}\tilde{Z}'X_{s*}\tilde{\beta}_s$$

$$\tilde{y} = (\Theta_1 + \Theta_2 - \Theta_3)y - (\Theta_1 + \Theta_2 - \Theta_3)X_{s*}\tilde{\beta}_s$$

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

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

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

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:

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

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

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

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

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

Intercept = $\tilde{\gamma}_N + \tilde{\alpha}_T$

The calculation of the covariance matrix is as follows:

$$\text{Var} [\tilde{y}] = \hat{\sigma}_e^2 (\Delta_N^{-1} - \Sigma_1 + \Sigma_2) + (\Theta_1 + \Theta_2 - \Theta_3) \text{Var} [\tilde{\beta}_s] (\Theta_1 + \Theta_2 - \Theta_3)'$$
where
\[
\Sigma_1 = \Delta_N^{-1} \mathbf{A} Q^{-1} \mathbf{A} \Delta_N^{-1} \mathbf{A} Q^{-1} \mathbf{A} \Delta_N^{-1}
\]
\[
\Sigma_2 = \Delta_N^{-1} \mathbf{A} Q^{-1} \Delta_T Q^{-1} \mathbf{A} \Delta_N
\]
\[
\text{Var}[\hat{\alpha}] = \delta_e^2 \left( Q^{-1} \mathbf{Z} \mathbf{Z} Q^{-1} \right) + \left( Q^{-1} \mathbf{Z} \mathbf{X}_s \right) \text{Var} \left[ \hat{\beta}_s \right] \left( \mathbf{X}_s \mathbf{Z} Q^{-1} \right)
\]
\[
\text{Cov} \left[ \hat{\alpha}, \hat{\gamma} \right] = \delta_e^2 \Delta_N^{-1} \left[ \mathbf{A} Q^{-1} \Delta_T - \mathbf{A} Q^{-1} \mathbf{A} \Delta_N^{-1} \mathbf{A} \right] Q^{-1} + \left( \Theta_1 + \Theta_2 - \Theta_3 \right) \text{Var} \left[ \hat{\beta}_s \right] \left( \mathbf{X}_s \mathbf{Z} Q^{-1} \right)
\]
\[
\text{Cov} \left[ \hat{\gamma}, \hat{\beta} \right] = \left( \Theta_1 + \Theta_2 - \Theta_3 \right) \text{Var} \left[ \hat{\beta}_s \right]
\]
\[
\text{Cov} \left[ \hat{\alpha}, \hat{\beta} \right] = \left( Q^{-1} \mathbf{Z} \mathbf{X}_s \right) \text{Var} \left[ \hat{\beta}_s \right]
\]

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:

\[
\text{Cov} \left( D_i^C, D_j^C \right) = \text{Cov} \left( \hat{\gamma}_i, \hat{\gamma}_j \right) + \text{Cov} \left( \hat{\gamma}_i, \hat{\alpha}_T \right) + \text{Cov} \left( \hat{\gamma}_j, \hat{\alpha}_T \right) + \text{Var} \left( \hat{\alpha}_T \right)
\]
\[
\text{Cov} \left( D_i^T, D_u^T \right) = \text{Cov} \left( \hat{\alpha}_t, \hat{\alpha}_u \right) - \text{Cov} \left( \hat{\alpha}_t, \hat{\alpha}_T \right) - \text{Cov} \left( \hat{\alpha}_u, \hat{\alpha}_T \right) + \text{Var} \left( \hat{\alpha}_T \right)
\]
\[
\text{Cov} \left( D_i^C, D_i^T \right) = \text{Cov} \left( \hat{\gamma}_i, \hat{\alpha}_t \right) + \text{Cov} \left( \hat{\gamma}_i, \hat{\alpha}_T \right) - \text{Cov} \left( \hat{\gamma}_i, \hat{\alpha}_u \right) - \text{Var} \left( \hat{\alpha}_T \right)
\]
\[
\text{Cov} \left( D_i^C, \hat{\beta} \right) = -\text{Cov} \left( \hat{\gamma}_i, \hat{\beta} \right) - \text{Cov} \left( \hat{\alpha}_T, \hat{\beta} \right)
\]
\[
\text{Cov} \left( D_i^T, \hat{\beta} \right) = -\text{Cov} \left( \hat{\alpha}_T, \hat{\beta} \right) + \text{Cov} \left( \hat{\alpha}_T, \hat{\alpha}_u \right)
\]

**Variance Covariance of Dummy Variables with Intercept**

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

\[
\text{Cov} \left( D_i^C, D_j^C \right) = \text{Cov} \left( \hat{\gamma}_i, \hat{\gamma}_j \right) + \text{Cov} \left( \hat{\gamma}_i, \hat{\gamma}_N \right) - \text{Cov} \left( \hat{\gamma}_j, \hat{\gamma}_N \right) + \text{Var} \left( \hat{\gamma}_N \right)
\]
\[
\text{Cov} \left( D_i^T, D_u^T \right) = \text{Cov} \left( \hat{\alpha}_t, \hat{\alpha}_u \right) - \text{Cov} \left( \hat{\alpha}_t, \hat{\alpha}_T \right) - \text{Cov} \left( \hat{\alpha}_u, \hat{\alpha}_T \right) + \text{Var} \left( \hat{\alpha}_T \right)
\]
\[
\text{Cov} \left( D_i^C, D_i^T \right) = \text{Cov} \left( \hat{\gamma}_i, \hat{\alpha}_t \right) - \text{Cov} \left( \hat{\gamma}_i, \hat{\alpha}_T \right) - \text{Cov} \left( \hat{\gamma}_i, \hat{\alpha}_u \right) + \text{Cov} \left( \hat{\gamma}_N, \hat{\alpha}_t \right)
\]
\[
\text{Cov} \left( D_i^C, \text{Intercept} \right) = \text{Cov} \left( \hat{\gamma}_i, \hat{\gamma}_N \right) + \text{Cov} \left( \hat{\gamma}_i, \hat{\alpha}_T \right) - \text{Cov} \left( \hat{\gamma}_j, \hat{\alpha}_T \right) - \text{Var} \left( \hat{\gamma}_N \right)
\]
\[
\text{Cov} \left( D_i^T, \text{Intercept} \right) = \text{Cov} \left( \hat{\alpha}_t, \hat{\alpha}_T \right) + \text{Cov} \left( \hat{\alpha}_t, \hat{\gamma}_N \right) - \text{Cov} \left( \hat{\alpha}_T, \hat{\gamma}_N \right) - \text{Var} \left( \hat{\alpha}_T \right)
\]
\[
\text{Cov} \left( D_i^C, \hat{\beta} \right) = -\text{Cov} \left( \hat{\gamma}_i, \hat{\beta} \right) - \text{Cov} \left( \hat{\gamma}_N, \hat{\beta} \right)
\]
\[
\text{Cov} \left( D_i^T, \hat{\beta} \right) = -\text{Cov} \left( \hat{\alpha}_T, \hat{\beta} \right) + \text{Cov} \left( \hat{\alpha}_T, \hat{\gamma}_N \right)
\]
\[
\text{Cov} \left( \text{Intercept}, \hat{\beta}_f \right) = -\text{Cov} \left( \hat{\alpha}_T, \hat{\beta} \right) - \text{Cov} \left( \hat{\gamma}_N, \hat{\beta} \right)
\]
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 1407 and “Two-Way Fixed-Effects Model” on page 1408. To eliminate the fixed effects, 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 \( X_{i,t} \) and \( y_{i,t} \), 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 \( \tilde{X}_{i,t} = X_{i,t} - X_{i,t-1} \) for regressors and \( \tilde{y}_{i,t} = y_{i,t} - y_{i,t-1} \) for the dependent variable.

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

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

The first-differenced estimator is

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

The resulting residual can be denoted as \( \tilde{u} = \tilde{y} - \tilde{\beta}_{fd} \tilde{X} \). The degree 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 \( y \) on the cross section means of \( \tilde{X}_s \). In other words, you fit the following regression:

\[
\bar{y}_{i} = \bar{x}_i \beta^{BG} + \eta_i
\]

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

\[
\bar{y}_{i,t} = \bar{x}_{i,t} \beta^{BT} + \zeta_t
\]

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 \( Z_0 = \text{diag}(J_{T_i}) \), \( P_0 = \text{diag}(\bar{J}_{T_i}) \), and \( Q_0 = \text{diag}(E_{T_i}) \), with \( \bar{J}_{T_i} = J_{T_i} / T_i \) and \( E_{T_i} = I_{T_i} - \bar{J}_{T_i} \). Define \( \bar{X}_s = Q_0 X_s \) and \( \bar{y} = Q_0 y \) and \( J \) as a vector of ones \( T_i \) long.

In the one-way model, estimation proceeds in a two-step fashion. First, you obtain estimates of the variance of the \( \sigma^2 \) and \( \sigma_v^2 \). 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 \( \theta \), and estimation proceeds via OLS on partial deviations from group means.

PROC PANEL gives the following options for variance component estimators.

Fuller and Battese’s 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

\[
R(\nu) = y'Z_0(Z_0Z_0)^{-1}Z_0'y \\
R(\beta|\nu) = ((\bar{X}_s'\bar{X}_s)^{-1}\bar{X}_s'\bar{y})(\bar{X}_s'\bar{y}) \\
R(\beta) = (X_s'X_s)^{-1}(X_s'y) \\
R(\nu|\beta) = R(\beta|\nu) + R(\nu) - R(\beta)
\]

The estimator of the error variance is given by

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

If the NOINT option is specified, the estimator is

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

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

\[
\hat{\sigma}_v^2 = (R(\nu|\beta) - (N - 1)\hat{\sigma}_\epsilon^2) / (M - \text{tr}(Z_0X_s(X_s'X_s)^{-1}X_s'Z_0))
\]

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’s 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

\[ q_1 = \tilde{u}'Q_0\tilde{u} \]
\[ q_2 = \tilde{u}'P_0\tilde{u} \]

where the residuals \( \tilde{u} \) are given by \( \tilde{u} = (I_M - jM'M/M)(y - X_s(\tilde{X}_s\tilde{X}_s)^{-1}\tilde{X}_s'y) \) if there is an intercept and by \( \tilde{u} = y - X_s(\tilde{X}_s\tilde{X}_s)^{-1}\tilde{X}_s'y \) if there is not. A vector of \( M \) ones is represented by \( j \).

Consider the expected values

\[ E(q_1) = (M - N - (K - 1))\sigma^2 \]
\[ E(q_2) = (N - 1 + \text{tr}[(X'_sQ_0X_s)^{-1}X'_sP_0X_s] - \text{tr}[(X'_sQ_0X_s)^{-1}X'_sJ'MX_s])\sigma^2 \]
\[ [M - \left( \sum_i T_i^2 / M \right)]\sigma^2 \]

where \( \hat{\sigma}_v^2 \) and \( \hat{\sigma}_\varepsilon^2 \) 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’s 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

\[ q_1 = \tilde{u}'_{OLS}Q_0\tilde{u}_{OLS} \]
\[ q_2 = \tilde{u}'_{OLS}P_0\tilde{u}_{OLS} \]

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

\[ E(\hat{q}_1) = E(\hat{u}'_{OLS}Q_0\hat{u}_{OLS}) = \delta_{11}\hat{\sigma}_v^2 + \delta_{12}\hat{\sigma}_\varepsilon^2 \]
\[ E(\hat{q}_2) = E(\hat{u}'_{OLS}P_0\hat{u}_{OLS}) = \delta_{21}\hat{\sigma}_v^2 + \delta_{22}\hat{\sigma}_e^2 \]

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

\[
\begin{align*}
\delta_{11} &= \text{tr} \left( (X'X)^{-1}X'Z_0'Z_0X \right) - \text{tr} \left( (X'X)^{-1}X'P_0X \right) (X'X)^{-1}X'Z_0'Z_0X \\
\delta_{12} &= M - N - K + \text{tr} \left( (X'X)^{-1}X'P_0X \right) \\
\delta_{21} &= M - 2\text{tr} \left( (X'X)^{-1}X'Z_0'Z_0X \right) + \text{tr} \left( (X'X)^{-1}X'P_0X \right) \\
\delta_{22} &= N - \text{tr} \left( (X'X)^{-1}X'P_0X \right)
\end{align*}
\]

where \( \text{tr}(\cdot) \) 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’s Method**

The Nerlove method for estimating variance components can be obtained by setting \( \text{VCOMP} = \text{NL} \). The Nerlove method (see Baltagi 1995, p. 17) is assured to give estimates of the variance components that are always positive. Furthermore, it is simple in contrast to the previous estimators.

If \( \gamma_i \) is the \( i \)th fixed effect, Nerlove’s method uses the variance of the fixed effects as the estimate of \( \hat{\sigma}_\gamma^2 \). You have \( \hat{\sigma}_\gamma^2 = \sum_{i=1}^N (\gamma_i - \bar{\gamma})^2 / \sum_{i=1}^N 1 \), where \( \bar{\gamma} \) is the mean fixed effect. The estimate of \( \sigma_e^2 \) 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 \( (\theta_i) \) as

\[
\theta_i = 1 - \sigma_e / w_i \\
w_i^2 = T_i \sigma_v^2 + \sigma_e^2
\]

where \( T_i \) is the \( i \)th cross section’s time observations.

Taking the \( \theta_i \), you form the partial deviations,

\[
\bar{y}_{it} = y_{it} - \theta_i \bar{y}_i, \\
\bar{x}_{it} = x_{it} - \theta_i \bar{x}_i.
\]

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

The random effects \( \beta \) 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

$$u_{it} = v_i + e_t + e_{it}$$

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 $X_*$ and $y_*$ 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 $D_t$ 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

$$Z = (Z_1, Z_2)$$

where $Z_1 = (D_1^\prime, D_2^\prime, \ldots, D_T^\prime)$ and $Z_2 = \text{diag}(D_1j_N, D_2j_N, \ldots, D_Tj_N)$.

The matrix $Z$ gives the dummy variable structure for the two-way model.

For notational ease, let

$$\Delta_N = Z_1^\prime Z_1, \Delta_T = Z_2^\prime Z_2, A = Z_2^\prime Z_1$$

$$\bar{Z} = Z_2 - Z_1 \Delta_N^{-1} A^\prime$$

$$\bar{\Delta}_1 = I_M - Z_1 \Delta_N^{-1} Z_1^\prime$$

$$\bar{\Delta}_2 = I_M - Z_2 \Delta_T^{-1} Z_2^\prime$$

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

$$P = (I_M - Z_1 \Delta_N^{-1} Z_1^\prime) - \bar{Z} Q^{-1} \bar{Z}^\prime$$

Fuller and Battese’s 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

$$\hat{\sigma}_e^2 = \bar{u}^\prime P \bar{u} / (M - T - N + 1 - (K - 1))$$

where $P$ 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

\[
E(q_N) = \sigma^2 \left[ M - T - K + 1 \right] + \sigma_v^2 \left[ M - T - \text{tr} \left( X_s \tilde{\Delta}_2 Z_1 \tilde{\Delta}_2 X_s \left( X_s \tilde{\Delta}_2 X_s \right)^{-1} \right) \right]
\]

\[
E(q_T) = \sigma^2 \left[ M - N - K + 1 \right] + \sigma_v^2 \left[ M - N - \text{tr} \left( X_s \tilde{\Delta}_1 Z_2 \tilde{\Delta}_1 X_s \left( X_s \tilde{\Delta}_1 X_s \right)^{-1} \right) \right]
\]

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 \( \sigma_v^2 \), the two equations are solved for \( \sigma^2 \) and \( \sigma_v^2 \).

**Wansbeek and Kapteyn’s 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

\[
\hat{\sigma}_e^2 = \hat{\mathbf{u}}^T \mathbf{P} \hat{\mathbf{u}} / (M - T - N + 1 - (K - 1))
\]

where the \( \hat{\mathbf{u}} \) are given by

\[
\hat{\mathbf{u}} = (\mathbf{I}_M - \mathbf{j}_M \mathbf{j}_M^T / M)(\mathbf{y}_s - \mathbf{X}_s \mathbf{s}(\mathbf{X}_s \mathbf{P} \mathbf{s})^{-1} \mathbf{X}_s \mathbf{s} / \mathbf{P} \mathbf{s}) \text{ if there is an intercept}
\]

and by

\[
\hat{\mathbf{u}} = (\mathbf{y}_s - \mathbf{X}_s \mathbf{s}(\mathbf{X}_s \mathbf{P} \mathbf{s})^{-1} \mathbf{X}_s \mathbf{s} / \mathbf{P} \mathbf{s}) \text{ if there is not.}
\]

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 \( \hat{\mathbf{u}} \), equating their expected values to the realized quadratic forms, and solving for the variance components.

Let

\[
q_N = \hat{\mathbf{u}}^T \mathbf{Z}_2 \tilde{\Delta}_2^{-1} \mathbf{Z}_2^T \hat{\mathbf{u}}
\]

\[
q_T = \hat{\mathbf{u}}^T \mathbf{Z}_1 \tilde{\Delta}_1^{-1} \mathbf{Z}_1^T \hat{\mathbf{u}}
\]

The expected values are

\[
E(q_N) = (T + k_N - (1 + k_0))\sigma^2 + (T - \frac{\lambda_1}{M})\sigma_v^2 + (M - \frac{\lambda_2}{M})\sigma_e^2
\]

\[
E(q_T) = (N + k_T - (1 + k_0))\sigma^2 + (M - \frac{\lambda_1}{M})\sigma_v^2 + (N - \frac{\lambda_2}{M})\sigma_e^2
\]

where

\[
k_0 = j_M \mathbf{s} \mathbf{s} (\mathbf{X}_s \mathbf{P} \mathbf{s})^{-1} \mathbf{X}_s \mathbf{s} j_M / M
\]
Chapter 20: The PANEL Procedure

\[ k_N = \text{tr}((X'_{s\delta}PX_{s\delta})^{-1}X'_{s\delta}Z_2\Delta_T^{-1}Z_2'X_{s\delta}) \]
\[ k_T = \text{tr}((X'_{s\delta}PX_{s\delta})^{-1}X'_{s\delta}Z_1\Delta_N^{-1}Z_1'X_{s\delta}) \]
\[ \lambda_1 = j'_MZ_1Z_1'j_M \]
\[ \lambda_2 = j'_MZ_2Z_2'j_M \]

The quadratic unbiased estimators for \( \sigma^2 \) and \( \sigma^2_e \) 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):

\[ E(q_N) = (T + k_N)\sigma^2 + T\sigma^2_v + M\sigma^2_e \]
\[ E(q_T) = (N + k_T)\sigma^2 + M\sigma^2_v + N\sigma^2_e \]

**Wallace and Hussain’s 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:

\[ q_e = \tilde{u}'_{OLS}P\tilde{u}_{OLS} \]
\[ q_N = \tilde{u}'_{OLS}Z_2\Delta_T^{-1}Z_2'\tilde{u}_{OLS} \]
\[ q_T = \tilde{u}'_{OLS}Z_1\Delta_N^{-1}Z_1'\tilde{u}_{OLS} \]

Taking expectations yields

\[ E(q_e) = E(\tilde{u}'_{OLS}P\tilde{u}_{OLS}) = \delta_{11}\sigma^2_e + \delta_{12}\sigma^2_v + \delta_{13}\sigma^2_e \]
\[ E(q_N) = E(\tilde{u}'_{OLS}Z_2\Delta_T^{-1}Z_2'\tilde{u}_{OLS}) = \delta_{21}\sigma^2_e + \delta_{22}\sigma^2_v + \delta_{23}\sigma^2_e \]
\[ E(q_T) = E(\tilde{u}'_{OLS}Z_1\Delta_N^{-1}Z_1'\tilde{u}_{OLS}) = \delta_{31}\sigma^2_e + \delta_{32}\sigma^2_v + \delta_{33}\sigma^2_e \]

where the \( \delta_{js} \) constants are defined by

\[ \delta_{11} = M - N - T + 1 - \text{tr} \left( X'PX(X'X)^{-1} \right) \]
\[ \delta_{12} = \text{tr} \left( X'Z_1Z_1'X(X'X)^{-1}(X'PX(X'X)^{-1}) \right) \]
\[ \delta_{13} = \text{tr} \left( X'Z_2Z_2'X(X'X)^{-1}(X'PX(X'X)^{-1}) \right) \]
\[ \delta_{21} = T - \text{tr} \left( X'Z_2\Delta_T^{-1}Z_2'X(X'X)^{-1} \right) \]
The PANEL procedure solves this system for the estimates \( \hat{\delta}_\epsilon, \hat{\delta}_\nu, \) and \( \hat{\delta}_e. \) Some of the estimated variance components can be negative. Negative components are set to zero and estimation proceeds.

**Nerlove's Method**

The Nerlove method for estimating variance components can be obtained with by setting VCOMP = NL. The estimator of the error variance is

\[
\hat{\sigma}_\epsilon^2 = \frac{\hat{u}'P\hat{u}}{M}
\]

The variance components for cross section and time effects are:

\[
\hat{\sigma}_\nu^2 = \sum_{i=1}^{N} \frac{(\gamma_i - \bar{\gamma})^2}{N-1} \quad \text{where } \gamma_i \text{ is the } i \text{th cross section effect}
\]

and

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

\[
y_{it} = y_i - \theta_1 y_{.i} - \theta_2 y_{.t} + \theta_3 y_{..}
\]

\[
x_{it} = x_i - \theta_1 x_{.i} - \theta_2 x_{.t} + \theta_3 x_{..}
\]

The \( \theta \) estimates are obtained from:

\[
\theta_1 = 1 - \frac{\sigma_e}{\sqrt{T \sigma_v^2 + \sigma_e^2}}
\]

\[
\theta_2 = 1 - \frac{\sigma_e}{\sqrt{N \sigma_v^2 + \sigma_e^2}}
\]

\[
\theta_3 = \theta_1 + \theta_2 + \frac{\sigma_e}{\sqrt{T \sigma_v^2 + N \sigma_v^2 + \sigma_e^2}} - 1;
\]

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 equation below:

\[
\Omega = \sigma^2_x I_M + \sigma^2_v Z_1 Z_1' + \sigma^2_e Z_2 Z_2'
\]

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

Wansbeek and Kapteyn show that the inverse of \( \Omega \) can be written as

\[
\sigma^2_e \Omega^{-1} = V - VZ_2 \tilde{P}^{-1} Z_2' V
\]

with the following:

\[
V = I_M - Z_1 \tilde{\Delta}_N^{-1} Z_1'
\]

\[
\tilde{P} = \tilde{\Delta}_T - A \tilde{\Delta}_N A'
\]

\[
\tilde{\Delta}_N = \Delta_N + \left( \frac{\sigma^2_v}{\sigma^2_e} \right) I_N
\]

\[
\tilde{\Delta}_T = \Delta_T + \left( \frac{\sigma^2_v}{\sigma^2_e} \right) I_T
\]

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.
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

$$
E(u_{it}^2) = \sigma_{ii} \text{(heteroscedasticity)}
$$

$$
E(u_{it}u_{jt}) = \sigma_{ij} \text{(contemporaneously correlated)}
$$

$$
u_{it} = \rho_i u_{i,t-1} + \epsilon_{it} \text{(autoregression)}
$$

where

$$
E(\epsilon_{it}) = 0
$$

$$
E(u_{i,t-1}\epsilon_{jt}) = 0
$$

$$
E(\epsilon_{it}\epsilon_{jt}) = \phi_{ij}
$$

$$
E(\epsilon_{it}\epsilon_{js}) = 0 \text{ for } s \neq t
$$

$$
E(u_{i0}) = 0
$$

$$
E(u_{i0}u_{j0}) = \sigma_{ij} = \phi_{ij}/(1 - \rho_i\rho_j)
$$

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

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

where

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

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

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

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

$$
\hat{\rho}_i = \left(\frac{\sum_{t=2}^{T}\hat{u}_{i,t}\hat{u}_{i,t-1}}{\sum_{t=2}^{T}\hat{u}_{i,t-1}^2}\right) \quad i = 1, 2, \ldots, N
$$
Finally, the autoregressive characteristic of the data is removed (asymptotically) by the usual transformation of taking weighted differences. That is, for \( i = 1, 2, \ldots, N \),

\[
y_{11} \sqrt{1 - \hat{\rho}_i^2} = \sum_{k=1}^{p} X_{1k} \sqrt{1 - \hat{\rho}_i^2} + u_{11} \sqrt{1 - \hat{\rho}_i^2}
\]

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

which is written

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

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

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

\[
\hat{u}^* = y^* - X^* \hat{\beta}_{OLS}
\]

from which the consistent estimator of \( \sigma_{ij} \) is calculated as follows:

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

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,

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

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

\[
\hat{\beta}_P = (X^* (\hat{\Phi}^{-1} \otimes I_T) X^*)^{-1} X^* (\hat{\Phi}^{-1} \otimes I_T) y^*
\]

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

\[
\text{Var}(\hat{\beta}_P) = (X' \mathbf{V}^{-1} X)^{-1}
\]
Standard Corrections
For the PARKS option, the first-order autocorrelation coefficient must be estimated for each cross section. Let $\rho$ 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:

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

where

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

and

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

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

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

where

$x_{it}' = (x_{it1}, \ldots, x_{itp})$ 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

$$y = X\beta + u$$

where

$$u = (a \otimes 1_T) + (1_N \otimes b) + e$$
\[ y = (y_{11}, \ldots, y_{1T}, y_{21}, \ldots, y_{NT})' \]

\[ X = (x_{11}, \ldots, x_{1T}, x_{21}, \ldots, x_{NT})' \]

\[ a = (a_1 \ldots a_N)' \]

\[ b = (b_1 \ldots b_T)' \]

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

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. \( x_{it} \) is a sequence of nonstochastic, known \( p \times 1 \) vectors in \( \mathcal{R}^p \) whose elements are uniformly bounded in \( \mathcal{R}^p \). The matrix \( X \) has a full column rank \( p \).

2. \( \beta \) is a \( p \times 1 \) constant vector of unknown parameters.

3. \( a \) is a vector of uncorrelated random variables such that \( E(a_i) = 0 \) and \( \text{var}(a_i) = \sigma_a^2 \), \( \sigma_a^2 > 0 \), \( i = 1, \ldots, N \).

4. \( b \) is a vector of uncorrelated random variables such that \( E(b_t) = 0 \) and \( \text{var}(b_t) = \sigma_b^2 \) where \( \sigma_b^2 > 0 \) and \( t = 1, \ldots, T \).

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

\[ e_{it} = \alpha_0 e_{i,t-1} + \alpha_1 e_{i,t-2} + \ldots + \alpha_m e_{i,t-m} \quad t = 1, \ldots, T; i = 1, \ldots, N \]

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

If assumptions 1–6 are satisfied, then

\[ E(y) = X\beta \]

and

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

where \( \Psi_T \) is a \( T \times T \) matrix with elements \( \psi_{ts} \) as follows:

\[ \text{Cov}(e_{it}e_{is}) = \begin{cases} \psi(|t - s|) & \text{if } |t - s| \leq m \\ 0 & \text{if } |t - s| > m \end{cases} \]
where \( \psi(k) = \alpha^2 \sum_{j=0}^{m-k} \alpha_j \alpha_{j+k} \) for \( k = |t-s| \). For the definition of \( I_N, I_T, J_N, \) and \( J_T, \) see the section “Fuller and Battese’s Method” on page 1415.

The covariance matrix, denoted by \( V \), can be written in the form

\[
V = \sigma_a^2 (I_N \otimes J_T) + \sigma_b^2 (J_N \otimes I_T) + \sum_{k=0}^{m} \psi(k) (I_N \otimes \Psi^{(k)}_T)
\]

where \( \Psi^{(0)}_T = I_T \), and, for \( k = 1, \ldots, m \), \( \Psi^{(k)}_T \) is a band matrix whose \( k \)th off-diagonal elements are 1’s and all other elements are 0’s.

Thus, the covariance matrix of the vector of observations \( y \) has the form

\[
\text{Var}(y) = \sum_{k=1}^{m+3} v_k V_k
\]

where

\[
\begin{align*}
 v_1 &= \sigma_a^2 \\
 v_2 &= \sigma_b^2 \\
 v_k &= \psi(k-3)k = 3, \ldots, m+3 \\
 V_1 &= I_N \otimes J_T \\
 V_2 &= J_N \otimes I_T \\
 V_k &= I_N \otimes \Psi^{(k-3)}_T k = 3, \ldots, m+3
\end{align*}
\]

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

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 \( \sum_{i=1}^{n} \delta_i v_i \).

The parameters \( v_i (i = 1, \ldots, n) \) are associated with a linear model \( E(y) = X \beta \) with covariance matrix \( \sum_{i=1}^{n} v_i V_i \) 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 Estimator**

For an example on dynamic panel estimation using GMM option, see “Example 20.6: The Cigarette Sales Data: Dynamic Panel Estimation with GMM” on page 1487.

Consider the case of the following general model:

\[
y_{it} = \sum_{i=1}^{n} \phi_i y_{i(t-1)} + \sum_{k=1}^{K} \beta_k x_{itk} + \gamma_i + \alpha_t + \epsilon_{it}
\]

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 \( E(x_{it}^P \epsilon_{is}) \neq 0 \) for \( s < t \).
and zero otherwise. The variable \( x_{it}^* \) is defined as strictly exogenous if \( E(x_{it}^* \epsilon_{is}) = 0 \) for all \( s \) and \( t \). The \( \gamma_i \) and \( \alpha_t \) 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:

\[
\Delta y_{it} = \phi \Delta y_{i(t-1)} + v_{it}
\]

where \( v_{it} = \epsilon_{it} - \epsilon_{it-1} \).

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,

\[
\Delta y_{i2} = \phi \Delta y_{i1} + v_{i2}
\]

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

For \( t = 3 \) you have,

\[
\Delta y_{i3} = \phi \Delta y_{i2} + v_{i3}
\]

Clearly, you have every reason to suspect that \( \text{Cov}(\epsilon_{i1}, v_{i3}) = 0 \). This condition forms one restriction.

For \( t = 4 \), both \( \text{Cov}(\epsilon_{i1}, v_{i4}) = 0 \) and \( \text{Cov}(\epsilon_{i2}, v_{i4}) = 0 \) must hold.

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

\[
Z_i = \begin{pmatrix}
  y_{i1} & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\
  0 & y_{i1} & y_{i2} & 0 & \cdots & 0 & 0 & 0 & 0 \\
  0 & 0 & 0 & y_{i1} & y_{i2} & y_{i3} & 0 & \cdots & 0 \\
  \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
  0 & 0 & 0 & 0 & 0 & 0 & y_{i1} & \cdots & y_{i(T-2)}
\end{pmatrix}
\]

Using the instrument matrix, you form the weighting matrix \( A_N \) as

\[
A_N = \left( \frac{1}{N} \sum_{i} Z'_i H_i Z_i \right)^{-1}
\]

The initial weighting matrix is

\[
H_i = \begin{pmatrix}
  2 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & 0 \\
  -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 \\
  0 & -1 & 2 & -1 & 0 & \cdots & 0 & 0 & 0 & 0 \\
  \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
  0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\
  0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2
\end{pmatrix}
\]
Note that the maximum size of the $H_i$ matrix is $T-2$. The origins of the initial weighting matrix are the expected error covariances. Notice that on the diagonals,

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

and off diagonals,

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

If you let the vector of lagged differences (in the series $y_{it}$) be denoted as $\Delta y$, and the dependent variable as $\Delta y_i$, then the optimal GMM estimator is

$$\phi = \left[\left(\sum_i \Delta y_i'Z_i\right)A_N\left(\sum_i Z_i'\Delta y_i\right)\right]^{-1}\left(\sum_i \Delta y_i'Z_i\right)A_N\left(\sum_i Z_i'\Delta y_i\right)$$

Using the estimate, $\hat{\phi}$, you can obtain estimates of the errors, $\hat{\epsilon}$, or the differences, $\hat{\nu}$. 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, $M = \sum_{i=1}^N (T_i - 2)$.

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

$$\sigma^2 \left[\left(\sum_i \Delta y_i'Z_i\right)A_N\left(\sum_i Z_i'\Delta y_i\right)\right]^{-1}$$

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

Instead of imposing the structure of the weighting, you form the $H_i$ matrix through the following:

$$H_i = \tilde{\nu}_i \tilde{\nu}_i'$$

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:

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

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\textsuperscript{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 a instrument or that an instrument be a regressor.

First, suppose that the variables are all correlated with the individual effects $y_i$. 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,

$$Z_i = \begin{bmatrix} x_{i1} \cdots x_{iT} & 0 & 0 & 0 & 0 \\ 0 & x_{i1} \cdots x_{iT} & 0 & 0 & 0 \\ 0 & 0 & x_{i1} \cdots x_{iT} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & x_{i1} \cdots x_{iT} \end{bmatrix}$$

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,

$$Z_i = \begin{bmatrix} x_{i1} & 0 & 0 & 0 & 0 \\ 0 & x_{i1} x_{i2} & 0 & 0 & 0 \\ 0 & 0 & x_{i1} \cdots x_{iT} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & x_{i1} \cdots x_{i(T-1)} \end{bmatrix}$$

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 $y_i$. 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

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

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

Because the variables are uncorrelated with $y_i$ and thus uncorrelated with the error term $\mu_{it}$ 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\textsuperscript{1}In this section, “correlated” means correlated with the individual effects and “uncorrelated” means uncorrelated with the individual effects.

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

$$Z_i^* = \begin{pmatrix}
Z_i & 0 & 0 & 0 & \cdots & 0 \\
0 & x_{i11}^p & x_{i12}^p & 0 & \cdots & 0 \\
0 & 0 & 0 & x_{i13}^p & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots \ & \vdots \\
0 & 0 & 0 & 0 & \cdots & x_{iT}^p
\end{pmatrix}$$

For exogenous variables $x_{i1t}$ Arellano and Bond (1991) use $E\left(T^{-1}\sum_{s=1}^{T} \mu_{i3}x_{i1t}^e\right) = 0$. PROC PANEL uses the same ones as the predetermined variables—that is, $E(\mu_{i2}x_{i1t}^e) = 0$ and $E(\mu_{it}x_{i1t}^e) = 0$ 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

$$Z_i^* = \begin{pmatrix}
Z_i & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & x_{i11}^p & x_{i12}^p & x_{i13}^e & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & x_{i3}^p & x_{i3}^e & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & \cdots & x_{iT}^p & x_{iT}^e
\end{pmatrix}$$

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:

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

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

$$Z_i^{\Delta} = \begin{pmatrix}
0 & 0 & 0 & \cdots & 0 \\
\Delta y_{i2} & 0 & \cdots & 0 \\
0 & 0 & \Delta y_{i3} & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots \ & \vdots \\
0 & 0 & 0 & \cdots & \Delta y_{i(T-1)}
\end{pmatrix}$$

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 $E(\mu_{it}l_{it}) = 0$ for $t = 1, \ldots, T$ for level equations and $E(\Delta \epsilon_{it}d_{it}) = 0$ for $t = 2, \ldots, T$ for differenced equations. The instrument matrix for the level and differenced equations are $Z_{li}$

---

²This happens when $\phi \to 1$ or as $\sigma_p^2/\sigma_e^2 \to \infty$. In this case, the lagged dependent variables $y_{i(t-t)}$ become weak instruments for the differenced variables $\Delta y_{it}$. 
and \( Z_{di} \), respectively, as follows:

\[
Z_{li} = \begin{pmatrix}
    l_{i1} & 0 & 0 & 0 & 0 \\
    0 & l_{i2} & 0 & 0 & 0 \\
    0 & 0 & l_{i3} & 0 & 0 \\
    \vdots & \vdots & \vdots & \vdots & \vdots \\
    0 & 0 & 0 & 0 & l_{iT}
\end{pmatrix}
\]

\[
Z_{di} = \begin{pmatrix}
    d_{i1} & 0 & 0 & 0 & 0 \\
    0 & d_{i2} & 0 & 0 & 0 \\
    0 & 0 & d_{i3} & 0 & 0 \\
    \vdots & \vdots & \vdots & \vdots & \vdots \\
    0 & 0 & 0 & 0 & d_{iT}
\end{pmatrix}
\]

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

\[
Z_i = \begin{pmatrix}
    Z_{di} & 0 & 0 & 0 & 0 \\
    0 & Z_{li}^e & Z_{li}^p & Z_{li}^y & 0
\end{pmatrix}
\]

where \( Z_{li}^e \) and \( Z_{li}^p \) 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

\[
A_N^* = \left( \frac{1}{N} \sum_i^N Z_i^* H_i^* Z_i^* \right)^{-1}
\]

where

\[
H_i^* = \begin{pmatrix}
    H_i & 0 & 0 & 0 & 0 \\
    0 & 1 & 0 & 0 & 0 \\
    0 & 0 & 1 & 0 & 0 \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & 0 & 0 & 1
\end{pmatrix}
\]

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

\[
\Delta y_{it} = \beta^* S_{it} + \alpha_t - \alpha_{t-1} + v_{it} \quad y_{it} = \beta^* S_{it} + \gamma_i + \alpha_t + \epsilon_{it}
\]

where \( S_{it} \) is the matrix of all explanatory variables—lagged endogenous, exogenous, and predetermined.

Let \( y_{it}^* \) be given by

\[
y_{it}^* = \begin{pmatrix}
    \Delta y_{it} \\
    y_{it}
\end{pmatrix} \quad \beta^* = \begin{pmatrix} \phi & \beta \end{pmatrix} \quad S_{it}^* = \begin{pmatrix} \Delta S_{it} \\
    S_{it}
\end{pmatrix} \quad e_{it}^* = \begin{pmatrix} v_i \\
    \mu_i = \epsilon_i + \gamma_i
\end{pmatrix}
\]

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

\[
\hat{\beta}_1^* = \left[ \left( \sum_i S_{it}^* Z_{i}^* \right) A_N^* \left( \sum_i Z_{i}^* S_{it}^* \right) \right]^{-1} \left( \sum_i S_{it}^* Z_{i}^* \right) A_N^* \left( \sum_i Z_{i}^* y_{it}^* \right)
\]
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,

\[
\sigma^2 = \frac{\hat{\mu}' \hat{\mu}}{M - p} = \frac{(y_{it} - \hat{\beta}_1^* S_{it})' (y_{it} - \hat{\beta}_1^* S_{it})}{M - p}
\]

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

\[
\left[ \left( \sum_i S_i^* Z_i^* \right) A_N^* \left( \sum_i Z_i^* S_i^* \right) \right]^{-1} \sigma^2
\]

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 \( H_i^* \) matrix is recalculated as

\[
H_{i,2}^* = \begin{pmatrix} \hat{\nu}_i \hat{\nu}_i' & 0 \\ 0 & \hat{\mu}_i \hat{\mu}_i' \end{pmatrix}
\]

And the weighting matrix becomes

\[
A_N^* (\hat{\beta}_1^*) = \left( \frac{1}{N} \sum_i Z_i^* H_{i,2}^* Z_i^* \right)^{-1}
\]

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

Let \( G \) denote a temporary matrix,

\[
G = \left[ \left( \sum_i S_i^* Z_i^* \right) A_N^* \left( \sum_i Z_i^* S_i^* \right) \right]^{-1} \left( \sum_i S_i^* Z_i^* \right) A_N^*
\]

The robust covariance estimate of \( \hat{\beta}_1^* \) is

\[
V^r (\hat{\beta}_1^*) = G A_N^{-1} (\hat{\beta}_1^*) G'
\]

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,

\[
\hat{\beta}_2^* = \left[ \left( \sum_i S_i^{**} Z_i^* \right) A_N^* (\hat{\beta}_1^*) \left( \sum_i Z_i^* S_i^* \right) \right]^{-1} \left( \sum_i S_i^{**} Z_i^* \right) A_N^* (\hat{\beta}_1^*) \left( \sum_i Z_i^* y_i^* \right)
\]

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

\[
V (\hat{\beta}_2^*) = \left[ \left( \sum_i S_i^{**} Z_i^* \right) A_N^* (\hat{\beta}_1^*) \left( \sum_i Z_i^* S_i^* \right) \right]^{-1}
\]
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 \( \hat{\beta}_2 \) around the true value \( \beta^* \) as a function of the one-step GMM estimator \( \hat{\beta}_1 \),

\[
\hat{\beta}_2 - \beta^* = \left[ \left( \sum_i S_i^T Z_i^T \right) A_N^* \left( \hat{\beta}_1 \right) \left( \sum_i Z_i^T S_i^T \right) \right]^{-1} \left( \sum_i S_i^T Z_i^T \right) A_N^* \left( \hat{\beta}_1 \right) \left( \sum_i Z_i^T e_i^* \right) \\
+ D_{\beta^* A_N^* (\beta^*)} \left( \hat{\beta}_2 - \beta^* \right) + O_p \left( N^{-1} \right)
\]

where \( D_{\beta^* A_N^* (\beta^*)} \) is the first derivative of \( \hat{\beta}_2 - \beta^* \) with regard to \( \beta^* \). The \( k \)th column of \( D \) is

\[
\{ D_{\beta^* A_N^* (\beta^*)} \}_k = \left[ \left( \sum_i S_i^T Z_i^T \right) A_N^* \left( \beta^* \right) \left( \sum_i Z_i^T S_i^T \right) \right]^{-1} \left( \sum_i S_i^T Z_i^T \right) A_N^* \left( \beta^* \right) \frac{\partial A_N^{-1} (\beta)}{\partial \beta_k} |_{\beta^*} \left( \sum_i Z_i^T S_i^T \right) \\
\times \left[ \left( \sum_i S_i^T Z_i^T \right) A_N^* \left( \beta^* \right) \left( \sum_i Z_i^T S_i^T \right) \right]^{-1} \left( \sum_i S_i^T Z_i^T \right) A_N^* \left( \beta^* \right) \left( \sum_i Z_i^T e_i^* \right) \\
- \left[ \left( \sum_i S_i^T Z_i^T \right) A_N^* \left( \beta^* \right) \left( \sum_i Z_i^T S_i^T \right) \right]^{-1} \left( \sum_i S_i^T Z_i^T \right) A_N^* \left( \beta^* \right) \frac{\partial A_N^{-1} (\beta)}{\partial \beta_k} |_{\beta^*} \left( \sum_i Z_i^T e_i^* \right)
\]

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

\[
\left[ \left( \sum_i S_i^T Z_i^T \right) A_N^* \left( \hat{\beta}_1 \right) \left( \sum_i Z_i^T S_i^T \right) \right]^{-1} \left( \sum_i S_i^T Z_i^T \right) A_N^* \left( \hat{\beta}_1 \right) \left( \sum_i Z_i^T \hat{e}_{i,2}^* \right) = 0
\]

and

\[
\frac{\partial A_N^{-1} (\beta)}{\partial \beta_k} |_{\beta^*} = -\frac{1}{N} \sum_i Z_i^T \begin{pmatrix} \Delta S_{i,k}^T v_i^* + v_i \Delta S_{i,k}^T 0 \\ 0 S_{i,k}^T \mu_i^* + \mu_i S_{i,k} \end{pmatrix} Z_i^*
\]

The first part vanishes and leaves

\[
\{ D_{\hat{\beta}_2^* A_N^* (\hat{\beta}_1)} \}_k = \frac{1}{N} \left[ \left( \sum_i S_i^T Z_i^T \right) A_N^* \left( \hat{\beta}_1 \right) \left( \sum_i Z_i^T S_i^T \right) \right]^{-1} \left( \sum_i S_i^T Z_i^T \right) A_N^* \left( \hat{\beta}_1 \right) \\
\left( \sum_i Z_i^T \begin{pmatrix} \Delta S_{i,k}^T v_i^* + v_i \Delta S_{i,k}^T 0 \\ 0 S_{i,k}^T \mu_i^* + \mu_i S_{i,k} \end{pmatrix} Z_i^* \right) A_N^* \left( \hat{\beta}_1 \right) \left( \sum_i Z_i^T \hat{e}_{i,2}^* \right)
\]
Plugging these into the Taylor expansion series yields

\[ V^c (\hat{\beta}^*_2) = V (\hat{\beta}^*_2) + D_{\hat{\beta}^*_2} A^*_N (\hat{\beta}^*_1) V (\hat{\beta}^*_2) + D'_{\hat{\beta}^*_2} A^*_N (\hat{\beta}^*_1) V' (\hat{\beta}^*_1) D'_{\hat{\beta}^*_2} A^*_N (\hat{\beta}^*_1) \]

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 \( \hat{A}^*_{N,k+1} \) denote the robust covariance matrix from iteration \( k \), which is used as the weighting matrix in iteration \( k + 1 \). Iterate until

\[
\max_{i,j \leq \text{dim}(\hat{A}^*_N)} \left| \frac{\hat{A}^*_{N,k+1}(i,j) - \hat{A}^*_{N,k}(i,j)}{\hat{A}^*_{N,k}(i,j)} \right| \leq \text{ATOL}
\]

or

\[
\max_{i \leq \text{dim}(\hat{\beta}^*_k)} \left| \frac{\hat{\beta}^*_{k+1}(i) - \hat{\beta}^*_k(i)}{|\hat{\beta}^*_k(i)|} \right| \leq \text{BTOL}
\]

where \( \text{ATOL} \) is the tolerance for convergence in the weighting matrix and \( \text{BTOL} \) is the tolerance for convergence in the parameter estimate matrix. The default convergence criteria is \( \text{BTOL} = 1E-8 \) for PROC PANEL.

**Specification Testing For Dynamic Panel**

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

\[
\left( \sum_i \eta'_i Z^*_i \right) A^*_N \left( \sum_i Z^*'_i \eta_i \right) \sigma^2
\]

where \( \eta_i \) 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

\[
\left( \sum_i \eta'_i Z^*_i \right) A^*_{N,2} \left( \sum_i Z^*'_i \eta_i \right)
\]

This definition of the Sargan test is used for all iterated estimations. The Sargan test is distributed as a \( \chi^2 \) 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 \( l \)th lag is significant.
Define $\omega_l$ as the lag of the differenced error, with zero padding for the missing values generated. Symbolically,

$$
\begin{pmatrix}
0 \\
\vdots \\
0 \\
\omega_{l,2} \\
\vdots \\
\omega_{l,T-1-l}
\end{pmatrix}
$$

You define the constant $k_0$ as

$$
k_0 (l) = \sum_i \omega_{l,i}^t v_i
$$

You next define the constant $k_1$ as

$$
k_1 (l) = \sum_i \omega_{l,i}^t H_l \omega_{l,i}
$$

Note that the choice of $H_l$ 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

$$
k_2 (l) = -2 \left( \sum_i \omega_{l,i}^t \Delta S_i \right) G \left( \sum_i \Delta S_i^t Z_i \right) A_{N,k} \left( \sum_i Z_i^t H_l \omega_{l,i} \right)
$$

The matrix $G$ is defined as

$$
G = \left[ \left( \sum_i \Delta S_i^t Z_i^* \right) A_{N,k}^* \left( \sum_i Z_i^* \Delta S_i^* \right) \right]^{-1}
$$

The constant $k_3$ is defined as

$$
k_3 (l) = \left( \sum_i \omega_{l,i}^t \Delta S_i \right) V (\beta^*) \left( \sum_i \Delta S_i^t \omega_{l,i} \right)
$$

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 at most there will be ten time observations for each variable that enters as an instrument. 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 \( \max(t - \text{MAXBAND}, 1) \) and the last instrument is \( t - 2 \). The first observation for a predetermined variable is \( \max(t - \text{MAXBAND}, 1) \) and the last is \( t - 1 \). The first and last observation for an exogenous variable is given in the following list:

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

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

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

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 \textsc{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 \( H \) matrices are zeroed, and the calculation is continued. Alternatively, you can elect to replace missing values and missing observations with zeros (\textsc{zero}), the overall mean of the series (\textsc{oam}), the cross-sectional mean (\textsc{csm}), or the time series mean (\textsc{tsm}).
Linear Hypothesis Testing

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

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

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

\[
F = \frac{(\hat{u}_u' \hat{u}_u - \hat{u}' \hat{u}) / J}{\hat{u}' \hat{u} / (M - K)}
\]

where

- \( \hat{u}_u \) is the residual vector from the restricted regression
- \( \hat{u} \) 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:

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

The advantage of calculating Wald in this manner is that it enables you to substitute a heteroscedasticity-corrected covariance matrix for the matrix \( V \). 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 \( \chi^2 \) 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.

Consider the simple linear model:

\[ y = X\beta + \epsilon \]

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

\[ \hat{\beta} = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + \epsilon) = \beta + (X'X)^{-1}X'\epsilon \]

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

\[ \lim_{n \to \infty} \left( \frac{1}{n} X' \epsilon \right) = 0 \]

If the regressors are nonrandom, then it is possible to write the variance of the estimated \( \beta \) as the following:

\[ \text{Var} \left( \beta - \hat{\beta} \right) = (X'X)^{-1}X'\Omega X(X'X)^{-1} \]

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

\[ L^{-1} y = L^{-1} X\beta + L^{-1} \epsilon \]

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

The resulting GLS \( \beta \) is

\[ \hat{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \]

---

3The 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 1407 and “Two-Way Fixed-Effects Model” on page 1408. 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.

Using the GLS $\beta$, you can write
\[
\hat{\beta} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y \\
= (X' \Omega^{-1} X)^{-1} X' (\Omega^{-1} X + \Omega^{-1} \epsilon) \\
= \beta + (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} \epsilon
\]

The resulting variance expression for the GLS estimator is
\[
\text{Var} (\beta - \hat{\beta}) = (X' \Omega^{-1} X)^{-1} \epsilon \epsilon' \Omega^{-1} X (X' \Omega^{-1} X)^{-1} \\
= (X' \Omega^{-1} X)^{-1} \epsilon \epsilon' \Omega^{-1} \Omega^{-1} X (X' \Omega^{-1} X)^{-1} \\
= (X' \Omega^{-1} X)^{-1}
\]

The difference in variance between the OLS estimator and the GLS estimator can be written as
\[
(X' X)^{-1} X \Omega X (X' X)^{-1} - (X' \Omega^{-1} X)^{-1}
\]

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 $\beta$. Estimation of the $\Omega$ 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 $\Omega$. On the contrary, it suffices to calculate an estimate of the middle expression. That is, you need an estimate of:

\[
\Lambda = X' \Omega X
\]

This matrix, $\Lambda$, is easier to estimate because its dimension is K. PROC PANEL provides the following classical HCCME estimators for $\Lambda$:

The matrix is approximated by:

- **HCCME=N0:**
  \[
  \hat{\sigma}^2 X' X
  \]
  This is the simple OLS estimator. If you do not specify the HCCME= option, PROC PANEL defaults to this estimator.

- **HCCME=0:**
  \[
  \sum_{i=1}^{N} \sum_{t=1}^{T_i} \hat{\epsilon}^2_{it} x_{it} x_{it}'
  \]
  where $N$ is the number of cross sections and $T_i$ is the number of observations in $i$th cross section. The $x_{it}'$ is from the $r$th observation in the $i$th cross section, constituting the $(\sum_{j=1}^{i-1} T_j + t)$th row of the
matrix X. If the CLUSTER option is specified, one extra term is added to the preceding equation so that the estimator of matrix Λ is

\[
\sum_{i=1}^{N} \sum_{t=1}^{T_i} \hat{e}_{it}^2 x_{it} x_{it}' + \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{s=1}^{t-1} \sum_{s=1}^{t} \hat{e}_{it} \hat{e}_{is} \left( x_{it} x_{is}' + x_{is} x_{it}' \right)
\]

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:**

\[
M \frac{N}{M - K} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \hat{e}_{it}^2 x_{it} x_{it}'
\]

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

\[
M \frac{N}{M - K} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \hat{e}_{it}^2 x_{it} x_{it}' + \frac{M}{M - K} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{s=1}^{t-1} \sum_{s=1}^{t} \hat{e}_{it} \hat{e}_{is} \left( x_{it} x_{is}' + x_{is} x_{it}' \right)
\]

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:**

\[
\sum_{i=1}^{N} \sum_{t=1}^{T_i} \frac{\hat{e}_{it}^2}{1 - \hat{h}_{it}} x_{it} x_{it}'
\]

The \( \hat{h}_{it} \) term is the \( \sum_{j=1}^{N} T_j \)th diagonal element of the hat matrix. The expression for \( \hat{h}_{it} \) is \( x_{it}' (X'X)^{-1} x_{it} \). The hat matrix attempts to adjust the estimates for the presence of influence or leverage points. With the CLUSTER option, the estimator becomes

\[
\sum_{i=1}^{N} \sum_{t=1}^{T_i} \frac{\hat{e}_{it}^2}{1 - \hat{h}_{it}} x_{it} x_{it}' + 2 \sum_{i=1}^{N} \sum_{t=1}^{T_i} t-1 \sum_{s=1}^{t} \frac{\hat{e}_{it} \hat{e}_{is}}{\sqrt{1 - \hat{h}_{it} \hat{h}_{is}}} \left( x_{it} x_{is}' + x_{is} x_{it}' \right)
\]

The formula is similar to the robust variance matrix estimator in Wooldridge (2002, p. 152) with the heteroskedasticity adjustment.

- **HCCME=3:**

\[
\sum_{i=1}^{N} \sum_{t=1}^{T_i} \frac{\hat{e}_{it}^2}{(1 - \hat{h}_{it})^2} x_{it} x_{it}'
\]

With the CLUSTER option, the estimator becomes

\[
\sum_{i=1}^{N} \sum_{t=1}^{T_i} \frac{\hat{e}_{it}^2}{(1 - \hat{h}_{it})^2} x_{it} x_{it}' + 2 \sum_{i=1}^{N} \sum_{t=1}^{T_i} t-1 \sum_{s=1}^{t} \frac{\hat{e}_{it} \hat{e}_{is}}{1 - \hat{h}_{it} \hat{h}_{is}} \left( x_{it} x_{is}' + x_{is} x_{it}' \right)
\]

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:

$$y_i = \alpha_i + X_{is} \hat{\beta} + \epsilon_i$$

The parameter estimates $\hat{\beta}$ and $\alpha_i$ are the result of least squares dummy variables (LSDV) or within estimator regressions, and $i$ is a vector of ones of length $T_i$. The estimate of the $ith$ cross section’s $X' \Omega X$ matrix (where the $s$ subscript indicates that no constant column has been suppressed to avoid confusion) is $X_i' \Omega X_i$. The estimate for the whole sample is:

$$X_s' \Omega X_s = \sum_{i=1}^{N} X_i' \Omega X_i$$

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 $\text{Var}(\beta)$ is used to drive the covariance matrices for the fixed effects and the LaGrange multiplier standard errors. Robust estimates of the covariance matrix for $\beta$ 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 $\Lambda$ in sandwich form is needed. With the HAC option, it is estimated as

$$\Lambda_{\text{HAC}} = a \sum_{i=1}^{N} \sum_{t=1}^{T_i} \hat{\epsilon}_it \hat{\epsilon}_it' + a \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{s=1}^{t-1} k\left(\frac{s-t}{b}\right) \hat{\epsilon}_is \hat{\epsilon}_is' \left(x_{it}x_{is}' + x_{is}x_{it}'\right)$$

, 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 20.2.

$^5$The HCCME=0 with CLUSTER option sets $k(.) = 1$. 
Table 20.2  Kernel Functions

<table>
<thead>
<tr>
<th>Kernel Name</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bartlett</td>
<td>( k(x) = \begin{cases} 1 -</td>
</tr>
<tr>
<td>Parzen</td>
<td>( k(x) = \begin{cases} 1 - 6x^2 + 6</td>
</tr>
<tr>
<td>Quadratic spectral</td>
<td>( k(x) = \frac{25}{12\pi^2 x^2} \left( \frac{\sin(6\pi x/5)}{6\pi x/5} - \cos \left( \frac{6\pi x}{5} \right) \right) )</td>
</tr>
<tr>
<td>Truncated</td>
<td>( k(x) = \begin{cases} 1 &amp;</td>
</tr>
<tr>
<td>Tukey-Hanning</td>
<td>( k(x) = \begin{cases} (1 + \cos(\pi x))/2 &amp;</td>
</tr>
</tbody>
</table>

When the \( \text{BANDWIDTH=ANDREWS} \) option is specified, the bandwidth parameter is estimated as shown in Table 20.3.

Table 20.3  Bandwidth Parameter Estimation

<table>
<thead>
<tr>
<th>Kernel Name</th>
<th>Bandwidth Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bartlett</td>
<td>( b = 1.1447(\alpha(1)T)^{1/3} )</td>
</tr>
<tr>
<td>Parzen</td>
<td>( b = 2.6614(\alpha(2)T)^{1/5} )</td>
</tr>
<tr>
<td>Quadratic spectral</td>
<td>( b = 1.3221(\alpha(2)T)^{1/5} )</td>
</tr>
<tr>
<td>Truncated</td>
<td>( b = 0.6611(\alpha(2)T)^{1/5} )</td>
</tr>
<tr>
<td>Tukey-Hanning</td>
<td>( b = 1.7462(\alpha(2)T)^{1/5} )</td>
</tr>
</tbody>
</table>

Let \( \{g_{ait}\} \) denote each series in \( \{g_{it} = \hat{\epsilon}_{it}\} \), and let \((\rho_a, \sigma_a^2)\) 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 \( \text{Var}(\epsilon_{ait}) = \sigma_a^2 \). The \( \alpha(1) \) and \( \alpha(2) \) are estimated with the following formulas:

\[
\alpha(1) = \frac{\sum_{a=1}^{k} 4\rho_a^2 \sigma_a^4}{\sum_{a=1}^{k} \sigma_a^4 (1 - \rho_a)^2} \quad \alpha(2) = \frac{\sum_{a=1}^{k} 4\rho_a^3 \sigma_a^4}{\sum_{a=1}^{k} \sigma_a^4 (1 - \rho_a)^3}
\]

When you specify \( \text{BANDWIDTH=NEWEYWEST94} \), according to Newey and West (1994) the bandwidth parameter is estimated as shown in Table 20.4.
### Table 20.4 Bandwidth Parameter Estimation

<table>
<thead>
<tr>
<th>Kernel Name</th>
<th>Bandwidth Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bartlett</td>
<td>$b = 1.1447({s_1/s_0}^2T)^{1/3}$</td>
</tr>
<tr>
<td>Parzen</td>
<td>$b = 2.6614({s_1/s_0}^2T)^{1/5}$</td>
</tr>
<tr>
<td>Quadratic spectral</td>
<td>$b = 1.3221({s_1/s_0}^2T)^{1/5}$</td>
</tr>
<tr>
<td>Truncated</td>
<td>$b = 0.6611({s_1/s_0}^2T)^{1/5}$</td>
</tr>
<tr>
<td>Tukey-Hanning</td>
<td>$b = 1.7462({s_1/s_0}^2T)^{1/5}$</td>
</tr>
</tbody>
</table>

The $s_1$ and $s_0$ are estimated with the following formulas:

$$s_1 = 2 \sum_{j=1}^{n} j \sigma_j$$  
$$s_0 = \sigma_0 + 2 \sum_{j=1}^{n} \sigma_j$$

where $n$ is the lag selection parameter and is determined by kernels, as listed in Table 20.5.

### Table 20.5 Lag Selection Parameter Estimation

<table>
<thead>
<tr>
<th>Kernel Name</th>
<th>Lag Selection Parameter</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bartlett</td>
<td>$n = c(T/100)^{2/9}$</td>
</tr>
<tr>
<td>Parzen</td>
<td>$n = c(T/100)^{4/25}$</td>
</tr>
<tr>
<td>Quadratic Spectral</td>
<td>$n = c(T/100)^{2/25}$</td>
</tr>
<tr>
<td>Truncated</td>
<td>$n = c(T/100)^{1/5}$</td>
</tr>
<tr>
<td>Tukey-Hanning</td>
<td>$n = c(T/100)^{1/5}$</td>
</tr>
</tbody>
</table>

The $c$ in Table 20.5 is specified by the C= option; by default, C=12.

The $\sigma_j$ is estimated with the equation

$$\sigma_j = T^{-1} \sum_{t=j+1}^{T} \left( \sum_{a=i}^{k} g_{at} \sum_{a=i}^{k} g_{at-j} \right), \ j = 0, ..., n$$

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

$$b = \begin{cases} 
\lceil \gamma T^r + c \rceil & \text{if BANDWIDTH=SAMPLESIZE(INT) option is specified} \\
\gamma T^r + c & \text{otherwise} 
\end{cases}$$

where $T$ is the sample size, $\lceil x \rceil$ is the largest integer less than or equal to $x$, and $\gamma$, $r$, 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 \( \Lambda_{HAC} \) is calculated by

\[
\Lambda_{HAC} = a \sum_{i=1}^{N} \left\{ \left( \sum_{t=1}^{T_i} w_{it} w_{it}' + \sum_{t=1}^{T_i} \sum_{s=1}^{t-1} k \left( \frac{t-s}{b} \right) (w_{it} w_{is}' + w_{is} w_{it}') \right) \left( I - A_i \right)^{-1} \left( I - A_i \right)^{-1}' \right\}
\]

### 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:

\[
R^2 = 1 - \frac{\hat{u}^\prime \hat{V}^{-1} \hat{u}}{y^\prime D \hat{V}^{-1} Dy}
\]

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

and \( D = I_M - jM_j^i_M \left( \frac{\hat{V}^{-1}}{jM_j^i_M} \right) \).

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

\[
R^2 = 1 - \frac{\hat{u}^\prime \hat{V}^{-1} \hat{u}}{y^\prime \hat{V}^{-1} y}
\]

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

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

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

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

$$m = (\hat{\beta}_b - \hat{\beta}_a)' (\hat{S}_b - \hat{S}_a)^{-1} (\hat{\beta}_b - \hat{\beta}_a)$$

where $\hat{S}_b$ and $\hat{S}_a$ are consistent estimates of the asymptotic covariance matrices of $\hat{\beta}_b$ and $\hat{\beta}_a$. Then $m$ is distributed $\chi^2$ 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 $\chi^2$ 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 $i$th residual from the OLS regression, then the Breusch-Pagan (BP) test for one-way random effects is

$$BP = \frac{NT}{2(T - 1)} \left[ \frac{\sum_{i=1}^{N} \left[ \sum_{t=1}^{T} \hat{u}_{it} \right]^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2} - 1 \right]$$

The BP test generalizes to the case of a two-way random-effects model (Greene 2000, p. 589). Specifically,

$$BP2 = \frac{NT}{2(T - 1)} \left[ \frac{\sum_{i=1}^{N} \left[ \sum_{t=1}^{T} \hat{u}_{it} \right]^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2} - 1 \right]^2 + \frac{NT}{2(N - 1)} \left[ \frac{\sum_{t=1}^{T} \left[ \sum_{i=1}^{N} \hat{u}_{it} \right]^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2} - 1 \right]^2$$

is distributed as a $\chi^2$ 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 BP 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 1995 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, $\epsilon \sim N(0, \sigma^2 I_n)$, 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 $LR = -2\log \left( (1 + qF/df_u)^{-NT/2} \right) \implies LR = qF + O\left(n^{-1}\right)$. Under $H_0$, $LR$ is asymptotically distributed as a chi-square with $q$ degrees of freedom.

\(^6\)For the unbalanced panel, the number of time series $T_i$ might be different. The number of observations needs to be redefined accordingly.
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 $\hat{\rho}_{ij}$.

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t=T_{ij}}^{T_{ij}} e_{it} e_{jt}}{\sqrt{\sum_{t=T_{ij}}^{T_{ij}} e_{it}^2} \sqrt{\sum_{t=T_{ij}}^{T_{ij}} e_{jt}^2}}$$

where $T_{ij}$ 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, $T_{ij} = 1$ and $\overline{T}_{ij} = T$. Let $T_{ij}$ denote the number of overlapped time periods ($T_{ij} = \overline{T}_{ij} - T_{ij} + 1$). Then the Breusch-Pagan LM test statistic can be constructed as

$$BP = \sum_{i=1}^{N} \sum_{j=i+1}^{N} T_{ij} \hat{\rho}_{ij}^2$$

When $N$ is fixed and $T_{ij} \to \infty$, $BP \to \chi^2 (N (N - 1) / 2)$. 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:

$$BP_s = \sqrt{\frac{1}{N (N - 1)}} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left( T_{ij} \hat{\rho}_{ij}^2 - 1 \right)$$

Under the null hypothesis, $BP_s \to \mathcal{N} (0, 1)$ 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 CDp Test

Pesaran (2004) proposes a cross-sectional dependence test that is also based on the pairwise correlation coefficients $\hat{\rho}_{ij}$.

$$CD = \sqrt{\frac{2}{N (N - 1)}} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \sqrt{T_{ij}} \hat{\rho}_{ij}$$

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 $T_{ij} \to \infty$, $\sqrt{T_{ij}} \hat{\rho}_{ij} \implies \mathcal{N} (0, 1)$. Therefore, for $N$ and $T_{ij}$ tending to infinity in any order, $CD \implies \mathcal{N} (0, 1)$.

To enhance the power against the alternative hypothesis of local dependence, Pesaran (2004) proposes the CDp test. Local dependence is defined with respect to a weight matrix, $W = (w_{ij})$. Therefore, the test can
Panel Data Unit Root Tests

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

\[ \Delta y_{it} = \delta y_{it-1} + \sum_{L=1}^{p_i} \theta_{iL} \Delta y_{it-L} + \alpha_{mi} d_{mt} + \varepsilon_{it} \quad m = 1, 2, 3 \]

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 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 \( i \). 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 1450. Denote the selected lag orders as \( \hat{p}_i \). The test is implemented in 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:

\[ \Delta y_{it} = \delta y_{it-1} + \sum_{L=1}^{\hat{p}_i} \hat{\theta}_{iL} \Delta y_{it-L} + \alpha_{mi} d_{mt} + \varepsilon_{it} \quad m = 1, 2, 3 \]

The two orthogonalized residuals are generated by the following two auxiliary regressions:

\[ \Delta y_{it} = \sum_{L=1}^{\hat{p}_i} \theta_{iL} \Delta y_{it-L} + \alpha_{mi} d_{mi} + \varepsilon_{it} \]

\[ y_{it-1} = \sum_{L=1}^{\hat{p}_i} \theta_{iL} \Delta y_{it-L} + \alpha_{mi} d_{mi} + v_{it-1} \]

The residuals are saved at \( \hat{\varepsilon}_{it} \) and \( \hat{v}_{it-1} \), respectively. To remove heteroscedasticity, the residuals \( \hat{\varepsilon}_{it} \) and \( \hat{v}_{it-1} \) are normalized by the regression standard error from the ADF regression. Denote the standard error as \( \hat{\sigma}_{\varepsilon i}^2 = \sum_{t=\hat{p}_i+2}^{T} (\hat{\varepsilon}_{it} - \hat{\delta}_i \hat{v}_{it-1})^2 / (T - \hat{p}_i - 1) \), and normalize residuals as

\[ \hat{\varepsilon}_{it} = \frac{\hat{\varepsilon}_{it}}{\hat{\sigma}_{\varepsilon i}}, \quad \hat{v}_{it-1} = \frac{\hat{v}_{it-1}}{\hat{\sigma}_{\varepsilon i}} \]
Step 2 The ratios of long-run to short-run standard deviations of $\Delta y_{it}$ are estimated. Denote the ratios and the long-run variances as $s_i$ and $\sigma y_i$, 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 1451. Then the ratios are estimated by $\hat{s}_i = \hat{\sigma} y_i / \hat{\sigma} s_i$. Let the average standard deviation ratio be $S_N = (1/N) \sum_{i=1}^{N} s_i$, and let its estimator be $\hat{S}_N = (1/N) \sum_{i=1}^{N} \hat{s}_i$.

Step 3 The panel test statistics are calculated. To calculate the $t$ statistic and the adjusted $t$ statistic, the following equation is estimated:

$$\tilde{\epsilon}_{it} = \delta \tilde{\epsilon}_{i,t-1} + \tilde{\epsilon}_{it}$$

The total number of observations is $N \tilde{T}$, with $\tilde{p} = \sum_{i=1}^{N} \hat{p}_i / N$, $\tilde{T} = T - \tilde{p} - 1$. The standard $t$ statistic for testing $\delta = 0$ is $t_\delta = \hat{\delta} / STD(\hat{\delta})$, with OLS estimator $\hat{\delta}$ and standard deviation $STD(\hat{\delta})$. However, the standard $t$ statistic diverges to negative infinity for models (2) and (3). Let $\hat{\sigma}_e$ be the root mean square error from the step 3 regression, and denote it as

$$\hat{\sigma}_e^2 = \left[ \frac{1}{N \tilde{T}} \sum_{i=1}^{N} \sum_{t=2+\tilde{p}_i}^{T} (\tilde{\epsilon}_{it} - \hat{\delta} \tilde{\epsilon}_{i,t-1})^2 \right]$$

Levin, Lin, and Chu (2002) propose the following adjusted $t$ statistic:

$$t_\delta^* = \frac{t_\delta - N \tilde{T} \hat{S}_N \hat{\sigma}_e^{-2} STD(\hat{\delta}) \mu^*_m \tilde{T}}{\sigma^*_m \tilde{T}}$$

The mean and standard deviation adjustments ($\mu^*_m, \sigma^*_m$) 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 (HQIC), 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 $2c k log [log(T_o)]$ with $c$ being a constant greater than 1.\footnote{In practice $c$ is set to 1, following the literature (Hannan and Quinn 1979; Hall 1994).} 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 \( \hat{\theta}_t \) 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 \( \Delta y_{it} \) is estimated by a HAC-type estimator. For model (1), given the lag truncation parameter \( \tilde{K} \) and kernel weights \( w_{KL} \), the formula is

\[
\hat{\sigma}^2_{\Delta y_{it}} = \frac{1}{T-1} \sum_{t=2}^{T} \Delta y_{it}^2 + 2 \sum_{L=1}^{\tilde{K}} w_{KL} \left[ \frac{1}{T-1} \sum_{t=2+L}^{T} \Delta y_{it} \Delta y_{it-\tilde{L}} \right]
\]

To achieve consistency, the lag truncation parameter must satisfy \( \tilde{K} / T \to 0 \) and \( \tilde{K} \to \infty \) as \( T \to \infty \). Levin, Lin, and Chu (2002) suggest \( \tilde{K} = \left[ 3.21 T^{1/3} \right] \). The weights \( w_{KL} \) 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 details about the kernel functions and Andrews (1991) data-driven bandwidth selection procedure, see the section “Heteroscedasticity- and Autocorrelation-Consistent Covariance Matrices” on page 1442 for details. Because Levin, Lin, and Chu (2002) truncated the bandwidth as an integer, when LLCBAND is specified as the BANDWIDTH option, it corresponds to BANDWIDTH = \( \left[ 3.21 T^{1/3} \right] + 1 \). Furthermore, kernel weights \( w_{KL} = k(L/((\tilde{K} + 1)) \) with kernel function \( k(\cdot) \).

For model (2), the series \( \Delta y_{it} \) is demeaned individual by individual first. Therefore, \( \Delta y_{it} \) is replaced by \( \Delta y_{it} - \bar{\Delta y}_{it} \), where \( \bar{\Delta y}_{it} \) is the mean of \( \Delta y_{it} \) 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 \( \Delta y_{it} \) on \{1, \( t \)\} for each individual and save the residual \( \widetilde{\Delta y}_{it} \), and then replace \( \Delta y_{it} \) 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 \( \theta_t \) denote the time-specific aggregate effects; then the data generating process (DGP) becomes

\[
\Delta y_{it} = \delta y_{i,t-1} + \sum_{L=1}^{p_t} \theta_{iL} \Delta y_{it-\tilde{L}} + \alpha_{mi} d_{mt} + \theta_t + \varepsilon_{it} \quad m = 4, 5
\]

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 but 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 \( \theta_t \) 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:

$$\Delta y_{it} = \beta_i y_{it-1} + \sum_{j=1}^{p_i} \rho_{ij} \Delta y_{i,t-j} + \alpha_{mi} d_{mt} + \varepsilon_{it} \quad m = 1, 2, 3$$

Unlike the DGP in LLC, $\beta_i$ 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,

$$H_1 : \beta_i < 0 \quad \text{for } i = 1, \ldots, N_1, \quad \beta_i = 0 \quad \text{for } i = N_1 + 1, \ldots, N$$

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 $t\text{-bar}_N$, is formed as a simple average of the individual $t$ statistics for testing the null hypothesis of $\beta_i = 0$. If $t_i$ ($p_i, \rho_i$) is the standard $t$ statistic, then

$$t\text{-bar}_N = \frac{1}{N} \sum_{i=1}^{N} t_i(p_i, \rho_i)$$

If $T \to \infty$, then for each $i$ the $t$ statistic (without time trend) converges to the Dickey-Fuller distribution, $\eta_i$, defined by

$$\eta_i = \frac{1}{2} \left\{ W_i(1)^2 - 1 \right\} - W_i(1) \int_0^1 W_i(u) du$$

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

$$Z_{t\text{bar}}(p, \rho) = \frac{\sqrt{N} \{ t\text{-bar}_N - E(\eta) \}}{\sqrt{\text{Var}(\eta)}} \implies \mathcal{N}(0, 1)$$
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

$$W_{t \text{bar}}(p, \rho) = \frac{\sqrt{N} \{ t \text{-bar}_{NT} - \sum_{i=1}^{N} E[t_{iT} (p_i, 0) | \beta_i = 0] / N \} \sqrt{\sum_{i=1}^{N} Var[t_{iT} (p_i, 0) | \beta_i = 0] / N}}$$

$$\Rightarrow \mathcal{N}(0, 1)$$

Im, Pesaran, and Shin (2003) simulate the values of $E[t_{iT} (p_i, 0) | \beta_i = 0]$ and $Var[t_{iT} (p_i, 0) | \beta_i = 0]$ 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). See the section “Lag Order Selection in the ADF Regression” on page 1450 for details.

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 ($\beta_i = 0$),

$$\tilde{t} \text{-bar}_{NT} = \sum_{i=1}^{N} \tilde{t}_{iT} / N = \sum_{i=1}^{N} \left[ \hat{\beta}_{iT} (y_{i,-1}' M_T y_{i,-1})^{1/2} / \tilde{\sigma}_{iT} \right] / N$$

where $\hat{\beta}_{iT}$ is the OLS estimator of $\beta_i$ (unrestricted model), $\tau_T = (1, 1, \ldots, 1)'$, $M_T = I_T - \tau_T (\tau_T' \tau_T)^{-1} \tau_T'$, and $y_{i,-1} = (y_{i0}, y_{i1}, \ldots, y_{iT-1})'$. Under the null hypothesis, the standardized $\tilde{t}$-bar statistic converges to a standard normal variate,

$$Z_{\tilde{t} \text{bar}} = \frac{\sqrt{N} \{ \tilde{t} \text{-bar}_{NT} - E(\tilde{t}_{iT}) \}}{\sqrt{\text{Var}(\tilde{t}_{iT})}} \Rightarrow \mathcal{N}(0, 1)$$

where $E(\tilde{t}_{iT})$ and $\text{Var}(\tilde{t}_{iT})$ 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 $t$-bar$_{NT}$ and $\tilde{t}$-bar$_{NT}$ are simulated.

Similar as in section “Levin, Lin, and Chu (2002)” on page 1449, 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. See section “Cross-Sectional Dependence via Time-Specific Aggregate Effects” on page 1451 for details.

**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

$$p_i = F(G_i)$$
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):

\[ \text{Chi-Square} = -2 \sum_{i=1}^{N} \ln(p_i) \]

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

\[ P_m = FI = \sum_{i=1}^{N} \left( -2 \ln(p_i) - 2 \right) / 2 \sqrt{N} = - \sum_{i=1}^{N} \left( \ln(p_i) + 1 \right) / \sqrt{N} \]

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,

\[ Z = \sum_{i=1}^{N} \Phi^{-1}(p_i) \]

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,

\[ L^* = \sqrt{k} L = \sqrt{k} \sum_{i=1}^{N} \ln \left( \frac{p_i}{1-p_i} \right) \]

where \( k = 3 \left( 5N + 4 \right) / \left( \pi^2 N \left( 5N + 2 \right) \right) \). When \( T_i \to \infty \) and \( N \) is fixed, \( L^* \Rightarrow \mathcal{t}_{5N+4} \). 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 1449, 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. See section “Cross-Sectional Dependence via Time-Specific Aggregate Effects” on page 1451 for details.

---

^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.
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

\[
\Delta y_{it} = \delta^* (y_{i,t-1} - y_{i1}) + v_{it}
\]

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:

\[
\Delta y_{it}^* = \sqrt{\frac{T - t}{T - t + 1}} \left[ \Delta y_{it} - (\Delta y_{i,t+1} + \ldots + \Delta y_{iT}) / (T - t) \right]
\]

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,

\[
UB = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} \Delta y_{it}^* \tilde{y}_{i,t-1} / \sigma_i^2}{\sqrt{\sum_{i=1}^{N} \sum_{t=2}^{T} \tilde{y}_{i,t-1}^2 / \sigma_i^2}}
\]

where \( \sigma_i^2 = E (\Delta y_{it} - \beta_i)^2 \). When \( \sigma_i^2 \) is unknown, it can be estimated as

\[
\hat{\sigma}_i^2 = \frac{1}{T} \sum_{t=2}^{T} (\Delta y_{it} - \bar{y}_{i,t+1} / (T - 1))^2 / (T - 2)
\]

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). 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 (\( \mu_i = 0, \beta_i = 0 \)):

\[
\Delta y_{it} = \sum_{L=1}^{p_i} \theta_i L \Delta y_{i,t-L} + \mu_i + \epsilon_{it}
\]

\[\text{See the section “Levin, Lin, and Chu (2002)” on page 1449 for details. The only difference is the standard error estimate } \hat{\sigma}_{ei}^2. \text{ Breitung suggests using } T - p_i - 2 \text{ instead of } T - p_i - 1 \text{ as in LLC to normalize the standard error.} \]
where $\hat{p}_t$ is a consistent estimator of the true lag order $p_t$ and can be estimated by the procedures listed in the section “Lag Order Selection in the ADF Regression” on page 1450. 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 $\left(\hat{\mu}_t, \hat{\theta}_{1L}\right)$ be the estimated coefficients. The prewhitened series can be obtained by

$$\Delta \hat{y}_{it} = \Delta y_{it} - \sum_{L=1}^{\hat{p}_t} \hat{\theta}_{iL} \Delta y_{it-L}$$

and

$$\hat{y}_{it} = y_{it} - \sum_{L=1}^{\hat{p}_t} \hat{\theta}_{iL} y_{it-L}$$

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 $N \to \infty$,

$$t_{OLS} = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} y_{i,t-1} \Delta y_{it}}{\hat{\sigma} \sqrt{\sum_{i=1}^{N} \sum_{t=2}^{T} y_{i,t-1}^2}} \implies N(0,1)$$

where $\hat{\sigma}^2 = \sum_{i=1}^{N} \sum_{t=2}^{T} \left(\Delta y_{it} - \hat{\delta} \hat{y}_{i,t-1}\right)^2 / N (T - 1)$ 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 $\Omega = E (v_t v_t')$ be a positive definite matrix with eigenvalues $\lambda_1 \geq \ldots \geq \lambda_N$. 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,

$$\Delta y_t = \delta y_{t-1} + v_t$$

The unknown covariance matrix $\Omega$ can be estimated by its sample counterpart,

$$\hat{\Omega} = \sum_{t=2}^{T} \left(\Delta y_{t} - \hat{\delta} \hat{y}_{t-1}\right) \left(\Delta y_{t} - \hat{\delta} \hat{y}_{t-1}\right)' / (T - 1)$$

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 $\nu_\Omega = \lim_{N \to \infty} \text{tr} (\Omega^2 / N) / (\text{tr} \Omega / N)^2$. The variance $\nu_\Omega$ can be consistently estimated by

$$\hat{\nu}_\delta = \left(\sum_{t=2}^{T} y_{t-1}' \hat{\Omega} y_{t-1}\right) / \left(\sum_{t=2}^{T} y_{t-1}' y_{t-1}\right)^2.$$ 

Thus the robust $t$ statistic can be calculated as

$$t_{rob} = \frac{\delta \hat{\nu}_\delta}{\hat{\nu}_\delta} = \frac{\sum_{t=2}^{T} y_{t-1}' \Delta y_t}{\sqrt{\sum_{t=2}^{T} y_{t-1}' \hat{\Omega} y_{t-1}}} \implies N(0,1)$$

11Breitung (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.
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

\[
\Delta \hat{z}_t = \delta \hat{z}_{t-1} + \epsilon_t
\]

The feasible GLS (FGLS) estimator of \( \delta \) 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 1449, 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. See section “Cross-Sectional Dependence via Time-Specific Aggregate Effects” on page 1451 for details.

**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,

  \[
  r_{it} = r_{i(t-1)} + u_{it} \quad i = 1, \ldots, N, \quad t = 1, \ldots, T
  \]

  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 \( \epsilon_{it} \) and \( u_{it} \) satisfy \( \epsilon_{it} \sim \text{iid}\mathcal{N}(0, \sigma^2_\epsilon) \), \( u_{it} \sim \text{iid}\mathcal{N}(0, \sigma^2_u) \) and are mutually independent.

  The null hypothesis of stationarity is \( H_0 : \sigma^2_u = 0 \) against the alternative random walk hypothesis \( H_1 : \sigma^2_u > 0 \).

  In matrix form, the models can be written as

  \[
  y_i = X_i \beta_i + \epsilon_i
  \]
where \( y'_i = (y_{i1}, \ldots, y_{iT}) \), \( e'_i = (e_{i1}, \ldots, e_{iT}) \) with \( e_{it} = \sum_{j=1}^{T} u_{ij} + \varepsilon_{it} \), and \( X_i = (u_{iT}, a_{iT}) \) with \( a_{iT} \) being a \( T \times 1 \) vector of ones, \( a'_i = (1, \ldots, T) \), and \( \beta'_i = (r_{io}, \hat{\beta}_i) \).

Let \( \hat{\varepsilon}_{ij} \) be the residuals from the regression of \( y_i \) on \( X_i \); then the LM statistic is

\[
LM = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{T} \frac{T}{\hat{\sigma}_\varepsilon^2} \sum_{t=1}^{T} S_{it}^2
\]

where \( S_{it} = \sum_{j=1}^{T} \hat{\varepsilon}_{ij} \) is the partial sum of the residuals and \( \hat{\sigma}_\varepsilon^2 \) is a consistent estimator of \( \sigma_\varepsilon^2 \) under the null hypothesis of stationarity. With some regularity conditions,

\[
LM \xrightarrow{D} E \left[ \int_{0}^{1} V^2 (r) \, dr \right]
\]

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),

\[
V (r) = \begin{cases} 
W (r) - rW (1) & \text{for model (1)} \\
W (r) + (2r - 3r^2) W (1) + 6r (r - 1) \int_{0}^{1} W (s) \, ds & \text{for model (2)}
\end{cases}
\]

The mean and variance of the random variable \( \int V^2 \) can be calculated by using the characteristic functions,

\[
\xi = E \left[ \int_{0}^{1} V^2 (r) \, dr \right] = \begin{cases} 
\frac{1}{6} & \text{for model (1)} \\
\frac{11}{15} & \text{for model (2)}
\end{cases}
\]

and

\[
\zeta^2 = var \left[ \int_{0}^{1} V^2 (r) \, dr \right] = \begin{cases} 
\frac{1}{12} \frac{1}{300} & \text{for model (1)} \\
\frac{1}{11} & \text{for model (2)}
\end{cases}
\]

The LM statistics can be standardized to obtain the standard normal limiting distribution,

\[
Z = \frac{\sqrt{N} (LM - \bar{\xi})}{\bar{\zeta}} \implies \mathcal{N} (0, 1)
\]

**Consistent Estimator of \( \sigma_\varepsilon^2 \)**

Hadri’s (2000) test can be applied to the general case of heteroscedasticity and serially correlated disturbance errors. Under homoscedasticity and serially uncorrelated errors, \( \sigma_\varepsilon^2 \) can be estimated as

\[
\hat{\sigma}_\varepsilon^2 = \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{\varepsilon}_{it}^2 / N (T - k)
\]

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 \( \sigma_{\varepsilon,i}^2 \) can be estimated by \( \hat{\sigma}_{\varepsilon,i}^2 = \sum_{t=1}^{T} \hat{\varepsilon}_{it}^2 / (T - k) \) for each individual \( i \) and the LM statistic needs to be modified to

\[
LM = \frac{1}{N} \sum_{i=1}^{N} \left( \frac{1}{T} \frac{T}{\hat{\sigma}_{\varepsilon,i}^2} \sum_{t=1}^{T} S_{it}^2 \right)
\]
To allow for temporal dependence over $t$, $\sigma^2$ has to be replaced by the long-run variance of $\epsilon_{it}$, which is defined as $\sigma^2 = \lim_{T \to \infty} T^{-1} (S_{iT}^2) / \tilde{N}$. A HAC estimator can be used to consistently estimate the long-run variance $\sigma^2$. For more information, see the section “Long-Run Variance Estimation” on page 1451.

Similar as in section “Levin, Lin, and Chu (2002)” on page 1449, 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. See section “Cross-Sectional Dependence via Time-Specific Aggregate Effects” on page 1451 for details.

**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,

$$y_{it} = \varphi y_{it-1} + \epsilon_{it}$$

Under the null hypothesis, $\varphi = 1$. For model (2), each series is a unit root process with a heterogeneous drift,

$$y_{it} = \alpha_i + \varphi y_{it-1} + \epsilon_{it}$$

Model (3) includes heterogeneous drifts and linear time trends,

$$y_{it} = \alpha_i + \beta_i t + \varphi y_{it-1} + \epsilon_{it}$$

Similar as in section “Levin, Lin, and Chu (2002)” on page 1449, 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. See section “Cross-Sectional Dependence via Time-Specific Aggregate Effects” on page 1451 for details.

Let $\hat{\varphi}$ be the OLS estimator of $\varphi$; then

$$\hat{\varphi} - 1 = \left[ \sum_{i=1}^{N} y'_{i,-1} Q_T y_{i,-1} \right]^{-1} \left[ \sum_{i=1}^{N} y'_{i,-1} Q_T \epsilon_i \right]$$

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 $Q_T = I_T - Z_T (Z'_T Z_T)^{-1} Z'_T$, 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 $N \to \infty$

$$\sqrt{N T} \left( (T-1)/2 \right) (\hat{\varphi} - 1) \overset{y_{i0}=0, H_0}{\sim} \mathcal{N}(0, 1)$$

As $T \to \infty$, it becomes $T \sqrt{N} (\hat{\varphi} - 1) \overset{H_0}{\Rightarrow} \mathcal{N}(0, 2)$.

When the drift is absent in model (2), $\alpha_i = 0$, under the null hypothesis, as $N \to \infty$

$$\sqrt{\frac{5N (T+1)^3 (T-1)}{3 (17T^2 - 20T + 17)}} \left( \hat{\varphi} - 1 + \frac{3}{(T + 1)} \right) \overset{\alpha_i=0, H_0}{\sim} \mathcal{N}(0, 1)$$
As \( T \to \infty \), \( \left( T \sqrt{N} (\hat{\phi} - 1) + 3 \sqrt{N} \right) / \sqrt{5175} \Rightarrow N(0, 1) \).

When the time trend is absent in model (3), \( \beta_t = 0 \), under the null hypothesis, as \( N \to \infty \)
\[
\sqrt{\frac{112N (T + 2)^3 (T - 2)}{15 (193T^2 - 728T + 1147)}} \left( \hat{\phi} - 1 + \frac{15}{2(T + 2)} \right) \beta_t = 0, H_0 \Rightarrow N(0, 1)
\]

When \( T \to \infty \), \( \left( T \sqrt{N} (\hat{\phi} - 1) + 7.5 \sqrt{N} \right) / \sqrt{2895/112} \Rightarrow N(0, 1) \).

---

**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 1445.

**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 \( H_1^0 : \sigma^2 = 0 \) (no cross-sectional effects) that is based on the pooled estimator. The alternative is the one-sided \( H_1^1 : \sigma^2 > 0 \). 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
\[
J = \sqrt{\frac{NT}{2(T - 1)} [d - 1]} \Rightarrow H_0 \Rightarrow N(0, 1)
\]

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
\[
S = \frac{J - E(J)}{\sqrt{\text{Var}(J)}} = \frac{d - E(d)}{\sqrt{\text{Var}(d)}} \Rightarrow N(0, 1)
\]

Let \( D = I_N \otimes J_T \), where \( J_T \) is the \( T \times T \) square matrix of 1s. The mean and variance can be calculated by the formulas
\[
E(d) = \text{Tr}(DM_Z)/(n - k)
\]
\[
\text{Var}(d) = 2\{(n - k)\text{Tr}(DM_Z)^2 - [\text{Tr}(DM_Z)]^2\}/((n - k)^2(n - k + 2))
\]
where \( \text{Tr} \) 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 \( \text{Tr}(DM_Z) \), let \( Z = (Z_1', Z_2', \ldots, Z_N)' \). Then
\[
\text{Tr}(DM_Z) = NT - \text{Tr}\left( J_T \sum_{i=1}^{N} \left[ Z_i \left( \sum_{j=1}^{N} Z_j Z_j^{-1} Z_i' \right) \right] \right)
\]
To test for $H_2^0 : \sigma_a^2 = 0$ (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

$$J2 = \sqrt{\frac{NT}{2(N-1)}} [d2 - 1] \sim \mathcal{N}(0,1)$$

$J2$ can be standardized by $D = J_N \otimes I_T$, and other parameters are unchanged. Therefore,

$$S2 = \frac{J2 - E(J2)}{\sqrt{\text{Var}(J2)}} = \frac{d2 - E(d2)}{\sqrt{\text{Var}(d2)}} \sim \mathcal{N}(0,1)$$

To test for $H_3^0 : \sigma_y^2 = 0, \sigma_a^2 = 0$ (no cross-sectional and time effects), the test statistic is $J3 = (J + J2)/\sqrt{2}$ and $D = \sqrt{n/(T-1)} (I_N \otimes J_T) / 2 + \sqrt{n/(N-1)} (J_N \otimes I_T) / 2$. To standardize, define $d3 = \sqrt{n/(T-1)} d/2 + \sqrt{n/(N-1)} (d2)/2$.

$$S3 = \frac{J3 - E(J3)}{\sqrt{\text{Var}(J3)}} = \frac{d3 - E(d3)}{\sqrt{\text{Var}(d3)}} \sim \mathcal{N}(0,1)$$

**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 $H_1^0$ and $H_2^0$, which coincides with the Honda (1985) UMP test. Baltagi, Chang, and Li (1992) extend the King and Wu (1997) test for $H_3^0$ as follows:

$$KW \equiv \sqrt{T-1} \frac{J}{\sqrt{N+T-2}} + \sqrt{N-1} \frac{J2}{\sqrt{N+T-2}} \sim \mathcal{N}(0,1)$$

For the standardization, use $D = I_N \otimes J_T + J_N \otimes I_T$. Define $d_{kw} = d + d2$; then

$$S_{kw} = \frac{KW - E(KW)}{\sqrt{\text{Var}(KW)}} = \frac{d_{kw} - E(d_{kw})}{\sqrt{\text{Var}(d_{kw})}} \sim \mathcal{N}(0,1)$$

**Gourieroux, Holly, and Monfort (1982) LM Test**

If one or both variance components ($\sigma_y^2$ and $\sigma_a^2$) 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 $H_3^0$, which is immune to the possible negative values of $J$ and $J2$. The test statistic is

$$\text{GHM} = \begin{cases} 
J2^2 + (J2)^2 & \text{if } J > 0, J2 > 0 \\
J2^2 & \text{if } J > 0, J2 \leq 0 \\
(J2)^2 & \text{if } J \leq 0, J2 > 0 \\
0 & \text{if } J \leq 0, J2 \leq 0 
\end{cases} \sim (1/4) \chi^2(0) + (1/2) \chi^2(1) + (1/4) \chi^2(2)$$

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:

\[ H_0^1 : \sigma_y^2 = 0, \theta = 0 \text{ for MA(1) with MA coefficient } \theta \text{ or } H_0^2 : \sigma_y^2 = 0, \rho = 0 \text{ for AR(1) with AR coefficient } \rho. \]

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

\[
BL91 = \frac{NT^2}{2(T-1)} \left[ A^2 - 4AB + 2TB^2 \right] \frac{H_{0}^{1,2}}{2} \xrightarrow{\chi^2} (2)
\]

where

\[
A = \frac{\sum_{i=1}^{N} \left( \sum_{t=1}^{T} \hat{u}_{it} \right)^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2} - 1, \quad B = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} \hat{u}_{it} \hat{u}_{i,t-1}}{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2}
\]

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 \( H_0 : \sigma_y^2 = 0 \), the errors \( u_{it} \) are serially uncorrelated. To test \( H_0 : \sigma_y^2 = 0 \), Wooldridge (2002) proposes to test for AR(1) serial correlation. The test statistic that he proposes is

\[
W = \frac{\sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \hat{u}_{it} \hat{u}_{is}}{\sum_{i=1}^{N} \left( \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \hat{u}_{it} \hat{u}_{is} \right)^{1/2}} \rightarrow \mathcal{N}(0, 1)
\]

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 1460, 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

\[
RS^*_p = \frac{NT^2 \left( B - \frac{A}{T} \right)^2}{(T-1)(1-\frac{2}{T})}
\]
Baltagi and Li (1991, 1995) derive the conventional RS test when the cross-sectional random effects is assumed to be absent:

\[ RS_{\rho} = \frac{NT^2B^2}{T - 1} \]

Symmetrically, to test for the cross-sectional random effects in the presence of serial correlation, the modified Rao’s score test statistic is

\[ RS_{\mu}^* = \frac{NT(A - 2B)^2}{2(T - 1)(1 - 2/T)} \]

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 \( \sigma_C^2 > 0 \), 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:

\[ RSO_{\mu}^* = \sqrt{\frac{NT}{2(T - 1)(1 - 2/T)}} (A - 2B) \rightarrow \mathcal{N}(0, 1) \]

**Baltagi and Li (1995) LM Test for First-Order Correlation under Fixed Effects**

Let \( \hat{u}_{it} \) be the residual from the fixed one-way model (FIXONE). The two-sided LM test statistic for testing a white noise component in a fixed one-way model (\( H_0^S : \theta = 0 \) or \( H_0^S : \rho = 0 \), given that \( \gamma_i \) are fixed effects) is

\[ BL95 = \frac{NT^2}{T - 1} \left( \frac{\sum_{i=1}^N \sum_{t=2}^T \hat{u}_{it} \hat{u}_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T \hat{u}_{it}^2} \right)^2 \]

The LM test statistic is asymptotically distributed as \( \chi^2(1) \) under the null hypothesis. The one-sided LM test with alternative hypothesis \( \rho > 0 \) is

\[ BL95_2 = \sqrt{\frac{NT^2}{T - 1} \frac{\sum_{i=1}^N \sum_{t=2}^T \hat{u}_{it} \hat{u}_{i,t-1}}{\sum_{i=1}^N \sum_{t=1}^T \hat{u}_{it}^2}} \]

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 \( H_0^D : \rho = 0 \) against the alternative \( H_1^D : 0 < |\rho| < 1 \) by the Durbin-Watson statistic based on residuals \( \hat{u}_{it} \) from the fixed one-way model (FIXONE):

\[ d_{\rho} = \frac{\sum_{i=1}^N \sum_{t=2}^T (\hat{u}_{it} - \hat{u}_{i,t-1})^2}{\sum_{i=1}^N \sum_{t=1}^T \hat{u}_{it}^2} \]

The test statistic \( d_{\rho} \) is a locally most powerful invariant test in the neighborhood of \( \rho = 0 \). Some of the upper and lower bounds are listed in Bhargava, Franzini, and Narendranathan (1982). For very large \( N \), to test against a positive correlation \( \rho > 0 \), you can simply test whether the test statistic \( d_{\rho} < 2 \).
Berenblut-Webb Statistic

Let \( \Delta \tilde{u}_{it} \) be the residuals from the first-difference estimation. Bhargava, Franzini, and Narendranathan (1982) suggest using the Berenblut-Webb statistic, which is a locally most powerful invariant test in the neighborhood of \( \rho = 1 \). The test statistic is

\[
g_\rho = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} \Delta \tilde{u}_{i,t}^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{u}_{i,t}^2}
\]

The upper and lower bounds are the same as for the Durbin-Watson statistic \( d_\rho \).

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 1463 and “Berenblut-Webb Statistic” on page 1464. Bhargava, Franzini, and Narendranathan (1982) also propose the \( R_\rho \) statistic to test the random walk null hypothesis \( \rho = 1 \) against the stationary alternative \(|\rho| < 1\). Let \( F^* = I_N \otimes F \), where \( F \) is a \((T - 1)(T - 1)\) symmetric matrix that has the following elements:

\[
F_{tt'} = \begin{cases} 
(T - t') t/T & \text{if } t' \geq t \\
-T & \text{if } t' < t \\
0 & \text{otherwise}
\end{cases} \quad (t, t' = 1, \ldots, T - 1)
\]

The test statistic is

\[
R_\rho = \frac{\Delta \tilde{U}' \Delta \tilde{U}}{\Delta \tilde{U}^* F^* \Delta \tilde{U}} = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} \Delta \tilde{u}_{i,t}^2}{\sum_{i=1}^{N} \sum_{t'=t}^{T-1}(T-t+1)\Delta \tilde{u}_{i,t}^2 + 2 \sum_{i=1}^{N} \sum_{t'=t}^{T-1}(T-t'+1)(t-1)\Delta \tilde{u}_{i,t} \Delta \tilde{u}_{i,t'}}/T
\]

The statistics \( R_\rho, g_\rho, \) and \( d_\rho \) can be used with the same bounds. They satisfy \( R_\rho \leq g_\rho \leq d_\rho \), 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 \( \phi \) 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 following data set shown in Figure 20.2 illustrates the preceding instance with the correct representation.

**Figure 20.2 Single Observation for Each Time Series**

<table>
<thead>
<tr>
<th>Obs</th>
<th>firm</th>
<th>year</th>
<th>production</th>
<th>cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1955</td>
<td>5.36598</td>
<td>1.14867</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1960</td>
<td>6.03787</td>
<td>1.45185</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1965</td>
<td>6.37673</td>
<td>1.52257</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1970</td>
<td>6.93245</td>
<td>1.76627</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1955</td>
<td>6.54535</td>
<td>1.35041</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1960</td>
<td>6.69827</td>
<td>1.71109</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1965</td>
<td>7.40245</td>
<td>2.09519</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>1970</td>
<td>7.82644</td>
<td>2.39480</td>
</tr>
</tbody>
</table>

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 20.3; this data set results in an error message due to multiple observations while using PROC PANEL.

**Figure 20.3 Multiple Observations for Each Time Series**

<table>
<thead>
<tr>
<th>Obs</th>
<th>firm</th>
<th>year</th>
<th>production</th>
<th>cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1955</td>
<td>5.36598</td>
<td>1.14867</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1955</td>
<td>6.37673</td>
<td>1.52257</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1960</td>
<td>6.03787</td>
<td>1.45185</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1970</td>
<td>6.93245</td>
<td>1.76627</td>
</tr>
<tr>
<td>5</td>
<td>2</td>
<td>1955</td>
<td>6.54535</td>
<td>1.35041</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>1960</td>
<td>6.69827</td>
<td>1.71109</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>1965</td>
<td>7.40245</td>
<td>2.09519</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>1970</td>
<td>7.82644</td>
<td>2.39480</td>
</tr>
</tbody>
</table>

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 a 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**

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 20.6  ODS Graphics Produced by PROC PANEL

<table>
<thead>
<tr>
<th>ODS Graph Name</th>
<th>Plot Description</th>
<th>Plots=Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>DiagnosticsPanel</td>
<td>All applicable plots listed below</td>
<td></td>
</tr>
<tr>
<td>ResidualPlot</td>
<td>Plot of the residuals</td>
<td>RESIDUAL, RESID</td>
</tr>
<tr>
<td>FitPlot</td>
<td>Predicted versus actual plot</td>
<td>FITPLOT</td>
</tr>
<tr>
<td>QQPlot</td>
<td>Plot of the quantiles of the residuals</td>
<td>QQ</td>
</tr>
<tr>
<td>ResidSurfacePlot</td>
<td>Surface plot of the residuals</td>
<td>RESIDSURFACE</td>
</tr>
<tr>
<td>PredSurfacePlot</td>
<td>Surface plot of the predicted values</td>
<td>PREDsurface</td>
</tr>
<tr>
<td>ActSurfacePlot</td>
<td>Surface plot of the actual values</td>
<td>ACTSURFACE</td>
</tr>
<tr>
<td>ResidStackPlot</td>
<td>Stack plot of the residuals</td>
<td>RESIDSTACK, RESSTACK</td>
</tr>
<tr>
<td>ResidHistogram</td>
<td>Plot of the histogram of residuals</td>
<td>RESIDUALHISTOGRAM, RESIDHISTOGRAM</td>
</tr>
</tbody>
</table>
**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, 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.

Say that your data set contains variables y, x1, x2, x3, and z2. The following statements result in a SAS data set:

```sas
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. In the FIXONE case, this is _FIXONE_ or _FIXONE-TIME_. If the model RANONE model is selected, the _METHOD_ variable is either _Ran1FB_, _Ran1WK_, _Ran1WH_, or _Ran1NL_, depending on the variance component estimators chosen.

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 20.7.

<table>
<thead>
<tr>
<th>ODS Table Name</th>
<th>Description</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ODS Tables Created by the MODEL Statement</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ModelDescription</td>
<td>Model description</td>
<td>Default</td>
</tr>
<tr>
<td>FitStatistics</td>
<td>Fit statistics</td>
<td>Default</td>
</tr>
<tr>
<td>FixedEffectsTest</td>
<td>$F$ test for no fixed effects</td>
<td>FIXONE,FIXTWO,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>FIXONETIME</td>
</tr>
<tr>
<td>ParameterEstimates</td>
<td>Parameter estimates</td>
<td>Default</td>
</tr>
<tr>
<td>CovB</td>
<td>Covariance of parameter estimates</td>
<td>COVB</td>
</tr>
<tr>
<td>CorrB</td>
<td>Correlations of parameter estimates</td>
<td>CORRB</td>
</tr>
<tr>
<td>VarianceComponents</td>
<td>Variance component estimates</td>
<td>RANONE, RANTWO,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DASILVA,</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RANONE, RANTWOC</td>
</tr>
<tr>
<td>RandomEffectsTest</td>
<td>Hausman test for random effects</td>
<td>RANONE, RANTWO</td>
</tr>
<tr>
<td></td>
<td></td>
<td>RANONE, RANTWO</td>
</tr>
<tr>
<td>AR1Estimates</td>
<td>First-order autoregressive parameter estimates</td>
<td>RHO(PARKS)</td>
</tr>
<tr>
<td>BFNTest</td>
<td>$R_p$ statistic for serial correlation</td>
<td>BFN</td>
</tr>
<tr>
<td>BL91Test</td>
<td>Baltagi and Li joint LM test</td>
<td>BL91</td>
</tr>
<tr>
<td>BL95Test</td>
<td>Baltagi and Li (1995) LM test</td>
<td>BL95</td>
</tr>
<tr>
<td>BreuschPaganTest</td>
<td>Breusch-Pagan one-way test</td>
<td>BP</td>
</tr>
<tr>
<td>BreuschPaganTest2</td>
<td>Breusch-Pagan two-way test</td>
<td>BP2</td>
</tr>
<tr>
<td>BSYTest</td>
<td>Bera, Sosa Escudero, and Yoon modified RS test</td>
<td>BSY</td>
</tr>
<tr>
<td>BWTest</td>
<td>Berenblut-Webb statistic for serial correlation</td>
<td>BW</td>
</tr>
<tr>
<td>DWTest</td>
<td>Durbin-Watson statistic for serial correlation</td>
<td>DW</td>
</tr>
</tbody>
</table>
Table 20.7 (continued)

<table>
<thead>
<tr>
<th>ODS Table Name</th>
<th>Description</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>GHMTest</td>
<td>Gourieroux, Holly, and Monfort two-way test</td>
<td>GHM</td>
</tr>
<tr>
<td>HondaTest</td>
<td>Honda one-way test</td>
<td>HONDA</td>
</tr>
<tr>
<td>HondaTest2</td>
<td>Honda two-way test</td>
<td>HONDA2</td>
</tr>
<tr>
<td>KingWuTest</td>
<td>King and Wu two-way test</td>
<td>KW</td>
</tr>
<tr>
<td>WOOLDTest</td>
<td>Wooldridge (2002) test for unobserved effects</td>
<td>WOOLDRIDGE02</td>
</tr>
<tr>
<td>CDTestResults</td>
<td>Cross-sectional dependence test</td>
<td>CDTEST</td>
</tr>
<tr>
<td>CDpTestResults</td>
<td>Local cross-sectional dependence test</td>
<td>CDTEST</td>
</tr>
<tr>
<td>Sargan</td>
<td>Sargan’s test for overidentification</td>
<td>GMM1, GMM2, ITGMM</td>
</tr>
<tr>
<td>ARTest</td>
<td>Autoregression test for the residuals</td>
<td>GMM1, GMM2, ITGMM</td>
</tr>
<tr>
<td>IterHist</td>
<td>Iteration history</td>
<td>ITPRINT(ITGMM)</td>
</tr>
<tr>
<td>ConvergenceStatus</td>
<td>Convergence status of iterated GMM estimator</td>
<td>ITGMM</td>
</tr>
<tr>
<td>EstimatedPhiMatrix</td>
<td>Estimated phi matrix</td>
<td>PARKS</td>
</tr>
<tr>
<td>EstimatedAutocovariances</td>
<td>Estimates of autocovariances</td>
<td>DASILVA</td>
</tr>
<tr>
<td>LLCResults</td>
<td>LLC panel unit root test</td>
<td>UROOTTEST</td>
</tr>
<tr>
<td>IPSResults</td>
<td>IPS panel unit root test</td>
<td>UROOTTEST</td>
</tr>
<tr>
<td>CTResults</td>
<td>Combination test for panel unit root</td>
<td>UROOTTEST</td>
</tr>
<tr>
<td>HadriResults</td>
<td>Hadri panel stationarity test</td>
<td>UROOTTEST</td>
</tr>
<tr>
<td>HTResults</td>
<td>Harris and Tzavalis panel unit root test</td>
<td>UROOTTEST</td>
</tr>
<tr>
<td>BRResults</td>
<td>Breitung panel unit root test</td>
<td>UROOTTEST</td>
</tr>
<tr>
<td>URootdetail</td>
<td>Panel unit root test intermediate results</td>
<td>UROOTTEST</td>
</tr>
<tr>
<td>PTestResults</td>
<td>Poolability test for panel data</td>
<td>POOLTEST</td>
</tr>
</tbody>
</table>

ODS Tables Created by the TEST Statement

TestResults Test results

Example: PANEL Procedure

Example 20.1: Analyzing Demand for Liquid Assets

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

```plaintext
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.

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

proc panel data=a;
  model d = y rd rt rs / fuller parks dasilva m=7;
  model t = y rd rt rs / parks;
  model s = y rd rt rs / parks;
  id state year;
run;
```

The income elasticities for liquid assets are greater than 1 except for the demand deposit income elasticity (0.692757) estimated by the Da Silva method. In Output 20.1.1, Output 20.1.2, and Output 20.1.3, 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 20.1.2, Output 20.1.3, and Output 20.1.5). Time deposits are also substitutes for S&L shares in the time deposit demand equation (see Output 20.1.4), while these liquid assets are independent of each other in Output 20.1.5 (insignificant coefficient estimate of RT, −0.02705). Demand deposits and time deposits appear to be weak complements in Output 20.1.3 and Output 20.1.4, while the cross elasticities between demand deposits and time deposits are not significant in Output 20.1.2 and Output 20.1.5.
Output 20.1.1  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

<table>
<thead>
<tr>
<th>Model Description</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
<td>RanTwo</td>
</tr>
<tr>
<td>Number of Cross Sections</td>
<td>7</td>
</tr>
<tr>
<td>Time Series Length</td>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fit Statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>0.0795</td>
</tr>
<tr>
<td>DFE</td>
<td>72</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0011</td>
</tr>
<tr>
<td>Root MSE</td>
<td>0.0332</td>
</tr>
<tr>
<td>R-Square</td>
<td>0.6786</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance Component Estimates</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance Component for Cross Sections</td>
<td>0.03427</td>
</tr>
<tr>
<td>Variance Component for Time Series</td>
<td>0.00026</td>
</tr>
<tr>
<td>Variance Component for Error</td>
<td>0.00111</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hausman Test for Random Effects</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>DF m Value Pr &gt; m</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5.51</td>
</tr>
<tr>
<td>0.2385</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td></td>
</tr>
<tr>
<td>DF</td>
<td></td>
</tr>
<tr>
<td>Estimate</td>
<td></td>
</tr>
<tr>
<td>Standard Error</td>
<td></td>
</tr>
<tr>
<td>t Value Pr &gt;</td>
<td>t</td>
</tr>
<tr>
<td>Intercept</td>
<td>1</td>
</tr>
<tr>
<td>-1.23606</td>
<td>0.7252</td>
</tr>
<tr>
<td>-1.70</td>
<td>0.0926</td>
</tr>
<tr>
<td>y</td>
<td>1</td>
</tr>
<tr>
<td>1.064058</td>
<td>0.1040</td>
</tr>
<tr>
<td>10.23</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Permanent Per Capita Personal Income</td>
<td></td>
</tr>
<tr>
<td>rd</td>
<td>1</td>
</tr>
<tr>
<td>-0.290943</td>
<td>0.0526</td>
</tr>
<tr>
<td>-5.53</td>
<td>&lt;.0001</td>
</tr>
<tr>
<td>Service Charge on Demand Deposits</td>
<td></td>
</tr>
<tr>
<td>rt</td>
<td>1</td>
</tr>
<tr>
<td>0.039388</td>
<td>0.0278</td>
</tr>
<tr>
<td>1.42</td>
<td>0.1603</td>
</tr>
<tr>
<td>Interest on Time Deposits</td>
<td></td>
</tr>
<tr>
<td>rs</td>
<td>1</td>
</tr>
<tr>
<td>-0.32662</td>
<td>0.1140</td>
</tr>
<tr>
<td>-2.86</td>
<td>0.0055</td>
</tr>
<tr>
<td>Interest on S &amp; L Association Shares</td>
<td></td>
</tr>
</tbody>
</table>

Output 20.1.2  Demand for Demand Deposits, Parks Method

The PANEL Procedure
Parks Method Estimation

Dependent Variable: d Per Capita Demand Deposits

<table>
<thead>
<tr>
<th>Model Description</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
<td>Parks</td>
</tr>
<tr>
<td>Number of Cross Sections</td>
<td>7</td>
</tr>
<tr>
<td>Time Series Length</td>
<td>11</td>
</tr>
</tbody>
</table>
Example 20.1: Analyzing Demand for Liquid Assets

Output 20.1.2 continued

| 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 20.1.3 Demand for Demand Deposits, DaSilva Method

The PANEL Procedure
Da Silva Method Estimation

Dependent Variable: d Per Capita Demand Deposits

<table>
<thead>
<tr>
<th>Model Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
</tr>
<tr>
<td>DaSilva</td>
</tr>
<tr>
<td>Number of Cross Sections</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>Time Series Length</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>Order of MA Error Process</td>
</tr>
<tr>
<td>7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fit Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
</tr>
<tr>
<td>21609.8923</td>
</tr>
<tr>
<td>DFE</td>
</tr>
<tr>
<td>72</td>
</tr>
<tr>
<td>MSE</td>
</tr>
<tr>
<td>300.1374</td>
</tr>
<tr>
<td>Root MSE</td>
</tr>
<tr>
<td>17.3245</td>
</tr>
<tr>
<td>R-Square</td>
</tr>
<tr>
<td>0.4995</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance Component Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance Component for Cross Sections</td>
</tr>
<tr>
<td>0.03063</td>
</tr>
<tr>
<td>Variance Component for Time Series</td>
</tr>
<tr>
<td>0.000148</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimates of Autocovariances</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lag</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
</tbody>
</table>
**Output 20.1.3 continued**

| 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 20.1.4 Demand for Time Deposits, Parks Method**

The PANEL Procedure

Parks Method Estimation

Dependent Variable: t Per Capita Time Deposits

**Model Description**

<table>
<thead>
<tr>
<th>Estimation Method</th>
<th>Parks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Cross Sections</td>
<td>7</td>
</tr>
<tr>
<td>Time Series Length</td>
<td>11</td>
</tr>
</tbody>
</table>

**Fit Statistics**

<table>
<thead>
<tr>
<th>SSE</th>
<th>34.5713</th>
<th>DFE</th>
<th>72</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.4802</td>
<td>Root MSE</td>
<td>0.6929</td>
</tr>
<tr>
<td>R-Square</td>
<td>0.9517</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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 20.1.5 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**

<table>
<thead>
<tr>
<th>Estimation Method</th>
<th>Parks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Cross Sections</td>
<td>7</td>
</tr>
<tr>
<td>Time Series Length</td>
<td>11</td>
</tr>
</tbody>
</table>
Example 20.2: 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:

\[
\ln(TC_{it}) = \alpha_N + \gamma_T + (\alpha_i - \alpha_N) + (\gamma_t - \gamma_T) + \beta_1 \ln(Q_{it}) + \beta_2 \ln(PF_{it}) + \beta_3 LF_{it} + \epsilon_{it}
\]

where the \(\alpha\) are the pure cross-sectional effects and \(\gamma\) are the time effects. The actual model speculated is highly nonlinear in the original variables. It would look like the following:

\[
TC_{it} = \exp(\alpha_i + \gamma_t + \beta_3 LF_{it} + \epsilon_{it}) Q_{it}^{\beta_1} P_{it}^{\beta_2}
\]

The data and preliminary SAS statements are:

```sas
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)";
```

... 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 20.2.1. The model is a two-way fixed-effects model. There are six cross sections and fifteen time observations.

Output 20.2.1  The Airline Cost Data—Model Description

The PANEL Procedure
Fixed Two Way Estimates

Dependent Variable: lC Log transformation of costs

<table>
<thead>
<tr>
<th>Model Description</th>
<th>FixTwo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
<td></td>
</tr>
<tr>
<td>Number of Cross Sections</td>
<td>6</td>
</tr>
<tr>
<td>Time Series Length</td>
<td>15</td>
</tr>
</tbody>
</table>

The R-square and degrees of freedom can be seen in Table 20.2.2. 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 20.2.2  The Airline Cost Data—Fit Statistics

<table>
<thead>
<tr>
<th>Fit Statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>0.1768</td>
</tr>
<tr>
<td>DFE</td>
<td>67</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0026</td>
</tr>
<tr>
<td>Root MSE</td>
<td>0.0514</td>
</tr>
<tr>
<td>R-Square</td>
<td>0.9984</td>
</tr>
</tbody>
</table>
The $F$ test for fixed effects is shown in Table 20.2.3. 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 20.2.3 The Airline Cost Data—Test for Fixed Effects

<table>
<thead>
<tr>
<th>F Test for No Fixed Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num DF</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>19</td>
</tr>
</tbody>
</table>

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 16th 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 20.2.4  The Airline Cost Data—Parameter Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>CS1</td>
<td>1</td>
<td>0.174237</td>
<td>0.0861</td>
<td>2.02</td>
<td>0.0470</td>
<td>Cross Sectional Effect 1</td>
</tr>
<tr>
<td>CS2</td>
<td>1</td>
<td>0.111412</td>
<td>0.0780</td>
<td>1.43</td>
<td>0.1576</td>
<td>Cross Sectional Effect 2</td>
</tr>
<tr>
<td>CS3</td>
<td>1</td>
<td>-0.04354</td>
<td>0.0519</td>
<td>-2.77</td>
<td>0.0073</td>
<td>Cross Sectional Effect 3</td>
</tr>
<tr>
<td>CS4</td>
<td>1</td>
<td>0.18019</td>
<td>0.0321</td>
<td>5.61</td>
<td>&lt;.0001</td>
<td>Cross Sectional Effect 4</td>
</tr>
<tr>
<td>CS5</td>
<td>1</td>
<td>-0.04671</td>
<td>0.0225</td>
<td>-2.08</td>
<td>0.0415</td>
<td>Cross Sectional Effect 5</td>
</tr>
<tr>
<td>TS1</td>
<td>1</td>
<td>-0.69286</td>
<td>0.3378</td>
<td>-2.05</td>
<td>0.0442</td>
<td>Time Series Effect 1</td>
</tr>
<tr>
<td>TS2</td>
<td>1</td>
<td>-0.63816</td>
<td>0.3321</td>
<td>-1.92</td>
<td>0.0589</td>
<td>Time Series Effect 2</td>
</tr>
<tr>
<td>TS3</td>
<td>1</td>
<td>-0.59554</td>
<td>0.3294</td>
<td>-1.81</td>
<td>0.0751</td>
<td>Time Series Effect 3</td>
</tr>
<tr>
<td>TS4</td>
<td>1</td>
<td>-0.54192</td>
<td>0.3189</td>
<td>-1.70</td>
<td>0.0939</td>
<td>Time Series Effect 4</td>
</tr>
<tr>
<td>TS5</td>
<td>1</td>
<td>-0.47288</td>
<td>0.2319</td>
<td>-2.04</td>
<td>0.0454</td>
<td>Time Series Effect 5</td>
</tr>
<tr>
<td>TS6</td>
<td>1</td>
<td>-0.42705</td>
<td>0.1884</td>
<td>-2.27</td>
<td>0.0267</td>
<td>Time Series Effect 6</td>
</tr>
<tr>
<td>TS7</td>
<td>1</td>
<td>-0.39586</td>
<td>0.1733</td>
<td>-2.28</td>
<td>0.0255</td>
<td>Time Series Effect 7</td>
</tr>
<tr>
<td>TS8</td>
<td>1</td>
<td>-0.33972</td>
<td>0.1501</td>
<td>-2.26</td>
<td>0.0269</td>
<td>Time Series Effect 8</td>
</tr>
<tr>
<td>TS9</td>
<td>1</td>
<td>-0.2718</td>
<td>0.1348</td>
<td>-2.02</td>
<td>0.0478</td>
<td>Time Series Effect 9</td>
</tr>
<tr>
<td>TS10</td>
<td>1</td>
<td>-0.22734</td>
<td>0.0763</td>
<td>-2.98</td>
<td>0.0040</td>
<td>Time Series Effect 10</td>
</tr>
<tr>
<td>TS11</td>
<td>1</td>
<td>-0.1118</td>
<td>0.0319</td>
<td>-3.50</td>
<td>0.0008</td>
<td>Time Series Effect 11</td>
</tr>
<tr>
<td>TS12</td>
<td>1</td>
<td>-0.03366</td>
<td>0.0429</td>
<td>-0.78</td>
<td>0.4354</td>
<td>Time Series Effect 12</td>
</tr>
<tr>
<td>TS13</td>
<td>1</td>
<td>-0.01775</td>
<td>0.0363</td>
<td>-0.49</td>
<td>0.6261</td>
<td>Time Series Effect 13</td>
</tr>
<tr>
<td>TS14</td>
<td>1</td>
<td>-0.01865</td>
<td>0.0305</td>
<td>-0.61</td>
<td>0.5430</td>
<td>Time Series Effect 14</td>
</tr>
<tr>
<td>Intercept</td>
<td>1</td>
<td>12.93834</td>
<td>2.2181</td>
<td>5.83</td>
<td>&lt;.0001</td>
<td>Intercept</td>
</tr>
<tr>
<td>IQ</td>
<td>1</td>
<td>0.817264</td>
<td>0.0318</td>
<td>25.66</td>
<td>&lt;.0001</td>
<td>Log transformation of quantity</td>
</tr>
<tr>
<td>IPF</td>
<td>1</td>
<td>0.168732</td>
<td>0.1635</td>
<td>1.03</td>
<td>0.3057</td>
<td>Log transformation of price of fuel</td>
</tr>
<tr>
<td>LF</td>
<td>1</td>
<td>-0.88267</td>
<td>0.2617</td>
<td>-3.37</td>
<td>0.0012</td>
<td>Load Factor (utilization index)</td>
</tr>
</tbody>
</table>

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 1465.

```plaintext
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 20.2.5 and Output 20.2.6.
Output 20.2.5  Diagnostic Panel 1

Fit Diagnostics for IC

Observations 90  MSE 0.002639  Model DF 67
Output 20.2.6  Diagnostic Panel 2

The UNPACK and ONLY options produce individual detail images of paneled plots. The graph shown in Output 20.2.7 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;
```

The UNPACK and ONLY options produce individual detail images of paneled plots. The graph shown in Output 20.2.7 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.
Example 20.3: The Airline Cost Data: Further Analysis

Using the same data as in Example 20.2, 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:

```plaintext
proc panel data=airline;
   id i t;
   model lC = lQ lPF LF / fixone;
run;
```

The preceding statements result in Output 20.3.1. The fit seems to have deteriorated somewhat. The SSE rises from 0.1768 to 0.2926.
You still reject poolability based on the $F$ test in Output 20.3.2 at all accepted levels of significance.

```
Output 20.3.2 The Airline Cost Data—Test for Fixed Effects

<table>
<thead>
<tr>
<th>F Test for No Fixed Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Num DF</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>
```

The parameters change somewhat dramatically as shown in Output 20.3.3. 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 20.3.3 The Airline Cost Data—Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>IQ</td>
</tr>
<tr>
<td>IPF</td>
</tr>
<tr>
<td>LF</td>
</tr>
</tbody>
</table>
```
Example 20.4: 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 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 is created by the PROC PANEL statements shown in Example 20.2. The PROC PANEL statements necessary to generate the estimates are as follows:

```sas
proc panel data=airline outest=estimates;
  id I T;
  RANONE: model lC = lQ lPF lF / ranone vcomp=fb;
  RANONEwk: model lC = lQ lPF lF / ranone vcomp=wk;
  RANONEwh: model lC = lQ lPF lF / ranone vcomp=wh;
  RANONEnl: model lC = lQ lPF lF / ranone vcomp=nl;
  RANTWO: model lC = lQ lPF lF / rantwo vcomp=fb;
  RANTWOwk: model lC = lQ lPF lF / rantwo vcomp=wk;
  RANTWOwh: model lC = lQ lPF lF / rantwo vcomp=wh;
  RANTWOnl: model lC = lQ lPF lF / rantwo vcomp=nl;
  POOLED: model lC = lQ lPF lF / pooled;
  BTWNG: model lC = lQ lPF lF / btwng;
  BTWNT: model lC = lQ lPF lF / btwnt;
run;

data table;
set estimates;
  VarCS = round(_VARCS_,.00001);
  VarTS = round(_VARTS_,.00001);
  VarErr = round(_VARERR_,.00001);
  Int = round(Intercept,.0001);
  lQ2 = round(lQ,.0001);
  lPF2 = round(lPF,.0001);
  lF2 = round(lF,.0001);
  if _n_ >= 9 then do;
    VarCS = . ;
    VarTS = . ;
  end;
  keep _MODEL_ _METHOD_ VarCS VarTS VarErr Int lQ2 lPF2 lF2;
run;
```

The parameter estimates and variance components for both models are reported in Output 20.4.1 and Output 20.4.2.
In the random-effects model, individual constant terms are viewed as randomly distributed across cross-sectional units and not as parametric shifts of the regression function, as in the fixed-effects model. This is appropriate when the sampled cross-sectional units are drawn from a large population. Clearly, in this example, the six airlines are a sample of all the airlines in the industry and not an exhaustive, or nearly exhaustive, list.

There are four ways of computing the variance components in the one-way random-effects model. The method by Fuller and Battese (1974) (FB), uses a “fitting of constants” methods to estimate them. The Wansbeek and Kapteyn (1989) (WK) method uses the true disturbances, while the Wallace and Hussain (WH) method uses ordinary least squares residuals.
Looking at the estimates of the variance components for cross section and error in Output 20.4.2, you see that equal variance components for error are computed for both FB and WK, while WH and NL are nearly equal. All four techniques produce different variance components for cross sections. These estimates are then used to estimate the values of the parameters in Output 20.4.1. All the parameters appear to have similar and equally plausible estimates. Both the index for output in revenue passenger miles (IQ) and fuel price (IPF) 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.

As in the one-way random-effects model, the variance components for error produced by the FB and WK methods are equal. 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 may be, they produce parameter estimates that are similar and plausible. As with the one-way effects model, the index for output (IQ) and fuel price (IPF) are small and positive. The load factor (LF) estimates are all negative and, with the exception of the estimate produced by the WH method, somewhat smaller than the estimates produced in the one-way model. During the time the data were collected, the Civil Aeronautics Board dissolved, so it is possible that the dummy variables are proxies for this dissolution. This 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. Since the model appears to have both cross-sectional and time series effects, the pooled model should not be used.

Finally, you examine the between groups estimators. For the between groups estimate, you are looking at each airline’s data averaged across time. You see in Output 20.4.1 that the between groups parameter estimates are radically different from all other parameter estimates. This 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 20.5: Using the FLATDATA Statement**

Sometimes the data can be found in compressed form, where each line consists of all observations for the dependent and independent variables for the cross section. To illustrate, suppose you have a data set with 20 cross sections where each cross section consists of observations for 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 represent other character and numeric variables that are constant across each cross section.

The observations for first five cross sections along with other variables are shown in Output 20.5.1. In this example, \( i \) represents the cross section. The time period is identified by the subscript on the \( Y \) and \( X \) variables; it ranges from 1 to 6.
Since the PANEL procedure cannot work directly with the data in compressed form, the FLATDATA statement can be used to transform the data. The OUT= option can be used to output transformed data to a data set.

```
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;
```

First, six observations for the uncompressed data set and results for the one-way fixed-effects model fitted are shown in Output 20.5.2 and Output 20.5.3.
Example 20.6: The Cigarette Sales Data: Dynamic Panel Estimation with GMM

Output 20.5.3 Estimation with the FLATDATA Statement

The PANEL Procedure
Fixed One Way Estimates

Dependent Variable: Y

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>X</td>
</tr>
</tbody>
</table>

Example 20.6: The Cigarette Sales Data: Dynamic Panel Estimation with GMM

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 Estimator” on page 1427. The data are a panel from 46 American states over the period 1963–92. For data description see: Baltagi and Levin (1992); Baltagi (1995). 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;
   ... more lines ...
The following statements sort the data by `STATE` and `YEAR` variables.

```sas
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:

```sas
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.

```sas
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. NOLEVELS options is specified to avoid use of level equations, as shown in the following statements:

```sas
proc panel data=cigar_lag;
    inst depvar;
    model lsales = lsales_1 lprice lndi lpimin
        / gmm2 nolevels maxband=5 noint;
    id state year;
run;
```
Output 20.6.1  Estimation with GMM

The PANEL Procedure
GMM: First Differences Transformation

Dependent Variable: Isales  Log cigarette sales in packs per capita

<table>
<thead>
<tr>
<th>Model Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
</tr>
<tr>
<td>Number of Cross Sections</td>
</tr>
<tr>
<td>Time Series Length</td>
</tr>
<tr>
<td>Estimate Stage</td>
</tr>
<tr>
<td>Maximum Number of Time Periods</td>
</tr>
</tbody>
</table>

Fit Statistics

<table>
<thead>
<tr>
<th></th>
<th>SSE</th>
<th>DFE</th>
<th>1284</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>1.7037</td>
<td>Root MSE</td>
<td>1.3053</td>
</tr>
</tbody>
</table>

Parameter Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Estimate</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>t</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Isales_1</td>
<td>1</td>
<td>0.572219</td>
<td>0.000981</td>
<td>583.51</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lprice</td>
<td>1</td>
<td>-0.23464</td>
<td>0.00306</td>
<td>-76.56</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indi</td>
<td>1</td>
<td>0.232673</td>
<td>0.000392</td>
<td>593.69</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lpmcin</td>
<td>1</td>
<td>-0.08299</td>
<td>0.00328</td>
<td>-25.29</td>
<td>&lt;.0001</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If the theory suggests that there are other valid instruments, PREDETERMINED, EXOGENOUS and CORRELATED options can also be used.

References


Subject Index

between estimators, 1414
  PANEL procedure, 1414
BY groups
  PANEL procedure, 1384
Da Silva method
  PANEL procedure, 1425
first-differenced methods, 1413
  one-way, 1413
  PANEL procedure, 1413
fixed-effects model
  one-way, 1407
Fuller Battese
  variance components, 1415
generalized least squares
  PANEL procedure, 1423
GMM in panel: Arellano and Bond’s estimator
  panel GMM, 1427
  GMM in panel: Arellano and Bond’s estimator, 1427
PANEL procedure
  between estimators, 1414
  BY groups, 1384
  Da Silva method, 1425
  first-differenced methods, 1413
  generalized least squares, 1423
  HAC, 1442
  HCCME =, 1438
  ID variables, 1386
  one-way fixed-effects model, 1407
  one-way random-effects model, 1415
  output data sets, 1466–1468
  output table names, 1469
  Parks method, 1423
  pooled estimator, 1414
  predicted values, 1401
  printed output, 1468
  R-square measure, 1445
  residuals, 1401
  specification tests, 1445
  two-way fixed-effects model, 1408
  two-way random-effects model, 1417
  Zellner’s two-stage method, 1424
ID variables
  PANEL procedure, 1386
linear hypothesis testing, 1437
Nerlove
  variance components, 1417
one-way
  first-differenced methods, 1413
  fixed-effects model, 1407
  random-effects model, 1415
  one-way fixed-effects model, 1407
  PANEL procedure, 1407
  one-way random-effects model
  PANEL procedure, 1415
output data sets
  PANEL procedure, 1466–1468
output table names
  PANEL procedure, 1469
two-way random-effects model, 1417
   PANEL procedure, 1417

variance components
   Fuller Battese, 1415
   Nerlove, 1417
   Wallace Hussain, 1416
   Wansbeek Kapteyn, 1416

Wallace Hussain
   variance components, 1416
Wansbeek Kapteyn
   variance components, 1416

Zellner's two-stage method
   PANEL procedure, 1424
Syntax Index

ALL option
   TEST statement (PANEL), 1402
ARTEST= option
   MODEL statement (PANEL), 1389
ATOL= option
   MODEL statement (PANEL), 1389
BANDOPT= option
   MODEL statement (PANEL), 1389
BASE = option
   PROC PANEL statement, 1385
BFN option
   MODEL statement (PANEL), 1389
BIASCORRECTED option
   MODEL statement (PANEL), 1389
BL91 option
   MODEL statement (PANEL), 1389
BL95 option
   MODEL statement (PANEL), 1390
BP option
   MODEL statement (PANEL), 1390
BSY option
   MODEL statement (PANEL), 1390
BTOL= option
   MODEL statement (PANEL), 1390
BTWNG option
   MODEL statement (PANEL), 1390
BW option
   MODEL statement (PANEL), 1390
BY statement
   PANEL procedure, 1384
CDTEST option
   MODEL statement (PANEL), 1390
CLASS statement
   PANEL procedure, 1385
CLUSTER option
   Model statement (PANEL), 1390
CORR option
   MODEL statement (PANEL), 1390
CORRB option
   MODEL statement (PANEL), 1390
CORROUT option
   PROC PANEL statement, 1383
COVB option
   MODEL statement (PANEL), 1390
COVOUT option
   PROC PANEL statement, 1383
DASILVA option
   MODEL statement (PANEL), 1390
DATA= option
   PROC PANEL statement, 1383
DW option
   MODEL statement (PANEL), 1391
FDONE option
   MODEL statement (PANEL), 1391
FDONETIME option
   MODEL statement (PANEL), 1391
FDTWO option
   MODEL statement (PANEL), 1391
FIXONE option
   MODEL statement (PANEL), 1391
FIXONETIME option
   MODEL statement (PANEL), 1391
FIXTWO option
   MODEL statement (PANEL), 1391
FLATDATA statement
   PANEL procedure, 1385
GHM option
   MODEL statement (PANEL), 1391
GINV= option
   MODEL statement (PANEL), 1391
GMM1 option
   MODEL statement (PANEL), 1391
GMM2 option
   MODEL statement (PANEL), 1391
HAC = option
   MODEL statement (PANEL), 1391
HCCME= option
   MODEL statement (PANEL), 1391
HONDA option
   MODEL statement (PANEL), 1391
Honda option
   MODEL statement (PANEL), 1391
ID statement
   PANEL procedure, 1386
INDID = option
   PROC PANEL statement, 1385
ITGMM option
   MODEL statement (PANEL), 1393
ITPRINT option
   MODEL statement (PANEL), 1393
KEEP = option
  PROC PANEL statement, 1385

KW option
  MODEL statement (PANEL), 1393

LAG statement
  PANEL procedure, 1388

LM option
  TEST statement (PANEL), 1402

LR option
  TEST statement (PANEL), 1402

M= option
  MODEL statement (PANEL), 1393

MAXBAND= option
  MODEL statement (PANEL), 1393

MAXITER= option
  MODEL statement (PANEL), 1393

MODEL statement
  PANEL procedure, 1388, 1389

NEWEYWEST= option
  MODEL statement (PANEL), 1394

NODIFFS option
  MODEL statement (PANEL), 1394

NOESTIM option
  MODEL statement (PANEL), 1394

NOINT option
  MODEL statement (PANEL), 1394

NOLEVELS option
  MODEL statement (PANEL), 1394

NOPRINT option
  MODEL statement (PANEL), 1394

OUT = option
  FlatData statement (PANEL), 1386

OUT= option
  OUTPUT statement (PANEL), 1401

OUTCORR option
  PROC PANEL statement, 1383

OUTCOV option
  PROC PANEL statement, 1383

OUTEST= option
  PROC PANEL statement, 1383, 1467

OUTPUT statement
  PANEL procedure, 1401
  PROC PANEL statement, 1466

OUTTRANS= option
  PROC PANEL statement, 1468

P= option
  OUTPUT statement (PANEL), 1401

PARKS option
  MODEL statement (PANEL), 1394

PHI option
  MODEL statement (PANEL), 1394

PLOTS option
  PROC PANEL statement, 1383

POOLED option
  MODEL statement (PANEL), 1394

POOLTEST option
  MODEL statement (PANEL), 1394

PREDICTED= option
  OUTPUT statement (PANEL), 1401

PRINTFIXED option
  MODEL statement (PANEL), 1395

PROC PANEL statement, 1383

R= option
  OUTPUT statement (PANEL), 1401

RANONE option
  MODEL statement (PANEL), 1395

RANTWO option
  MODEL statement (PANEL), 1395

RESIDUAL= option
  OUTPUT statement (PANEL), 1401

RHO option
  MODEL statement (PANEL), 1395

ROBUST option
  MODEL statement (PANEL), 1395

SINGULAR= option
  MODEL statement (PANEL), 1395

TIME option
  MODEL statement (PANEL), 1395

TSNAME = option
  PROC PANEL statement, 1385

VAR option
  MODEL statement (PANEL), 1390

VCOMP= option
  MODEL statement (PANEL), 1400

WALD option
  TEST statement (PANEL), 1402

WOOLDRIDGE02 option
  MODEL statement (PANEL), 1401