# Chapter 26
## The PANEL Procedure

<table>
<thead>
<tr>
<th>Contents</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview: PANEL Procedure</td>
<td>1802</td>
</tr>
<tr>
<td>Getting Started: PANEL Procedure</td>
<td>1804</td>
</tr>
<tr>
<td>Syntax: PANEL Procedure</td>
<td>1806</td>
</tr>
<tr>
<td>Functional Summary</td>
<td>1806</td>
</tr>
<tr>
<td>PROC PANEL Statement</td>
<td>1809</td>
</tr>
<tr>
<td>BY Statement</td>
<td>1811</td>
</tr>
<tr>
<td>CLASS Statement</td>
<td>1811</td>
</tr>
<tr>
<td>COMPARE Statement</td>
<td>1811</td>
</tr>
<tr>
<td>FLATDATA Statement</td>
<td>1814</td>
</tr>
<tr>
<td>ID Statement</td>
<td>1815</td>
</tr>
<tr>
<td>INSTRUMENTS Statement</td>
<td>1815</td>
</tr>
<tr>
<td>LAG, CLAG, SLAG, XLAG, and ZLAG Statements</td>
<td>1817</td>
</tr>
<tr>
<td>MODEL Statement</td>
<td>1818</td>
</tr>
<tr>
<td>OUTPUT Statement</td>
<td>1835</td>
</tr>
<tr>
<td>RESTRICT Statement</td>
<td>1835</td>
</tr>
<tr>
<td>TEST Statement</td>
<td>1836</td>
</tr>
<tr>
<td>Details: PANEL Procedure</td>
<td>1837</td>
</tr>
<tr>
<td>Specifying the Input Data</td>
<td>1837</td>
</tr>
<tr>
<td>Specifying the Regression Model</td>
<td>1838</td>
</tr>
<tr>
<td>Missing Values</td>
<td>1839</td>
</tr>
<tr>
<td>Unbalanced Data</td>
<td>1839</td>
</tr>
<tr>
<td>Common Notation</td>
<td>1839</td>
</tr>
<tr>
<td>Pooled Regression (POOLED Option)</td>
<td>1840</td>
</tr>
<tr>
<td>Between-Groups Regression (BTWNG and BTWNT Options)</td>
<td>1840</td>
</tr>
<tr>
<td>One-Way Fixed-Effects Model (FIXONE and FIXONETIME Options)</td>
<td>1840</td>
</tr>
<tr>
<td>Two-Way Fixed-Effects Model (FIXTWO Option)</td>
<td>1841</td>
</tr>
<tr>
<td>One-Way Fixed-Effects Model, First Differencing (FDONE and FDONETIME Options)</td>
<td>1842</td>
</tr>
<tr>
<td>Two-Way Fixed-Effects Model, First Differencing (FDTWO Option)</td>
<td>1843</td>
</tr>
<tr>
<td>One-Way Random-Effects Model (RANONE Option)</td>
<td>1843</td>
</tr>
<tr>
<td>Two-Way Random-Effects Model (RANTWO Option)</td>
<td>1846</td>
</tr>
<tr>
<td>Parks Method for Autoregressive Models (PARKS Option)</td>
<td>1849</td>
</tr>
<tr>
<td>Da Silva Method for Moving Average Models (DASILVA Option)</td>
<td>1851</td>
</tr>
<tr>
<td>Hausman-Taylor Estimation (HTAYLOR Option)</td>
<td>1854</td>
</tr>
<tr>
<td>Amemiya-MacCurdy Estimation (AMACURDY Option)</td>
<td>1855</td>
</tr>
<tr>
<td>Dynamic Panel Estimation (DYNDIFF and DYNYSYS Options)</td>
<td>1855</td>
</tr>
<tr>
<td>Restricted Estimation</td>
<td>1865</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 people, households, countries, firms, and so on. For example, in the case of survey data on household income, the panel is created by repeatedly surveying the same households over several years.

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

- one-way and two-way models
- fixed-effects, random-effects, and hybrid models
- autoregressive models
moving average models

dynamic panel models

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

Apart from the possible one-way or two-way nature of the effect, the other source of disparity between the possible specifications is the nature of the cross-sectional or time series effect. The models are referred to as fixed-effects models if the effects are nonrandom and as random-effects models otherwise.

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

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

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

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

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

- You can fit models for dynamic panel data by using the generalized method of moments (GMM).
- The Hausman-Taylor and Amemiya-MaCurdy estimators offer a compromise between fixed- and random-effects estimation in models where some variables are correlated with individual effects.
- The MODEL statement supports between and pooled estimation.
- The variance components for random-effects models can be calculated for both balanced and unbalanced panels by using the methods described by Fuller and Battese (1974); Wansbeek and Kapteyn (1989); Wallace and Hussain (1969); Nerlove (1971).
- The CLASS statement allows classification variables (and their interactions) directly in the analysis.
- The TEST statement performs Wald, Lagrange multiplier, and likelihood ratio tests.
- The RESTRICT statement specifies linear restrictions on the parameters.
- The FLATDATA statement processes data in compressed (wide) form.
Several methods that produce heteroscedasticity-consistent (HCCME) and heteroscedasticity- and autocorrelation-consistent (HAC) covariance matrices are supported, because the presence of heteroscedasticity and autocorrelation can result in inefficient and biased estimates of the covariance matrix in an OLS framework.

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

The LAG statement and related statements provide functionality for creating lagged variables from within the PANEL procedure. Using these statements is preferable to using the DATA step because creating lagged variables in a panel setting can prove difficult, often requiring multiple loops and careful consideration of missing values.

Working within the PANEL procedure makes the creation of lagged values easy. The LAG statement leaves missing values as is. Alternatively, missing values can be replaced with zeros, overall mean, time mean, or cross-sectional mean by using the ZLAG, XLAG, SLAG, or CLAG statement, respectively.

The OUTPUT statement enables you to output data and estimates for use in other analyses.

The COMPARE statement constructs tables that enable you to easily compare parameters across multiple models and estimators.

---

**Getting Started: PANEL Procedure**

The following DATA step creates the data set `Electricity` from the cost function data in Greene (1990). The variable `Production` is the log of output in millions of kilowatt-hours, and the variable `Cost` is the log of cost in millions of dollars.

```plaintext
data Electricity;
    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
;
```

Consider the model

\[ C_{it} = \beta_0 + \beta_1 P_{it} + v_i + e_{it} \]

for \( i = 1, \ldots, N \), \( t = 1, \ldots, T \)

where \( C_{it} \) represents cost, \( P_{it} \) represents production, \( v_i \) is the cross-sectional error component, and \( e_{it} \) is the error variance component.
The first step is to make sure the data are sorted by firms and years within firms:

```plaintext
proc sort data = Electricity;
  by firm year;
run;
```

If you assume that the cross-sectional effects are random, four possible estimators are available for the variance components. The `VCOMP=FB` option in the following statements uses the Fuller and Battese (1974) estimator to fit the model:

```plaintext
proc panel data = Electricity;
  id firm year;
  model cost = production / ranone vcomp = fb;
run;
```

The output of these statements is shown in Output 26.1.

**Figure 26.1 One-Way Random-Effects Estimation Results**

**The PANEL Procedure**

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

<table>
<thead>
<tr>
<th>Dependent Variable: cost</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model Description</td>
</tr>
<tr>
<td>Estimation Method</td>
</tr>
<tr>
<td>Number of Cross Sections</td>
</tr>
<tr>
<td>Time Series Length</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fit Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
</tr>
<tr>
<td>DFE</td>
</tr>
<tr>
<td>MSE</td>
</tr>
<tr>
<td>Root MSE</td>
</tr>
<tr>
<td>R-Square</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>Variance Component for Error</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hausman Test for Random Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficients DF m Value Pr &gt; m</td>
</tr>
<tr>
<td>1 1 9.08 0.0026</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>----------</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>production</td>
</tr>
</tbody>
</table>

Printed first is a report that provides the estimation method and various data counts. Fit statistics and variance components estimates are printed next. A Hausman specification test compares this model to its fixed-effects counterpart. Finally, the table of regression parameter estimates shows the estimates, standard errors, and $t$ tests.
Syntax: PANEL Procedure

The following statements are available in the PANEL procedure:

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

Functional Summary

The statements and options available in the PANEL procedure are summarized in Table 26.1.

<table>
<thead>
<tr>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Set Options</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Includes correlations in the OUTEST= data set</td>
<td>PROC PANEL</td>
<td>CORROUT</td>
</tr>
<tr>
<td>Includes covariances in the OUTEST= data set</td>
<td>PROC PANEL</td>
<td>COVOUT</td>
</tr>
<tr>
<td>Specifies the input data set</td>
<td>PROC PANEL</td>
<td>DATA=</td>
</tr>
<tr>
<td>Specifies variables to keep but not transform</td>
<td>FLATDATA</td>
<td>KEEP=</td>
</tr>
<tr>
<td>Specifies the output data set for the CLASS statement</td>
<td>CLASS</td>
<td>OUT=</td>
</tr>
<tr>
<td>Specifies the output data set</td>
<td>FLATDATA</td>
<td>OUT=</td>
</tr>
<tr>
<td>Specifies the name of an output SAS data set</td>
<td>OUTPUT</td>
<td>OUT=</td>
</tr>
<tr>
<td>Writes parameter estimates to an output data set</td>
<td>PROC PANEL</td>
<td>OUTTEST=</td>
</tr>
<tr>
<td>Writes the transformed series to an output data set</td>
<td>PROC PANEL</td>
<td>OUTTRANS=</td>
</tr>
<tr>
<td>Requests that the procedure produce graphics via the Output Delivery System</td>
<td>PROC PANEL</td>
<td>PLOTS</td>
</tr>
</tbody>
</table>

**Declaring the Role of Variables**

- Specifies BY-group processing: BY
- Specifies the classification variables: CLASS
- Converts the data to uncompressed form: FLATDATA
### Table 26.1 continued

<table>
<thead>
<tr>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specifies the cross-sectional and time ID variables</td>
<td>ID</td>
<td></td>
</tr>
<tr>
<td>Declares instrumental variables</td>
<td>INSTRUMENTS</td>
<td></td>
</tr>
</tbody>
</table>

#### Lag Generation
Speifies output data set for lags whose missing values are replaced by the cross-sectional mean | CLAG | OUT= |
Speifes output data set for lags that leave missing values unchanged | LAG | OUT= |
Speifes output data set for lags whose missing values are replaced by the time period mean | SLAG | OUT= |
Speifes output data set for lags whose missing values are replaced by the overall mean | XLAG | OUT= |
Speifes output data set for lags whose missing values are replaced by zero | ZLAG | OUT= |

#### Printing Control Options
Prints correlations of the estimates | MODEL | CORRB |
Prints covariances of the estimates | MODEL | COVB |
Suppresses printed output | MODEL | NOPRINT |
Requests that the procedure produce graphics via the Output Delivery System | MODEL | PLOTS |
Prints fixed effects | MODEL | PRINTFIXED |
Performs tests of linear hypotheses | TEST | |

#### Model Estimation Options
Specifies the Amemiya-MaCurdy model | MODEL | AMACURDY |
Requests the $R_p$ statistic for serial correlation under fixed effects | MODEL | BFN |
Requests the Baltagi and Li joint Lagrange multiplier (LM) test for serial correlation and random cross-sectional effects | MODEL | BL91 |
Requests the Baltagi and Li LM test for first-order correlation under fixed effects | MODEL | BL95 |
Requests the Breusch-Pagan test for one-way random effects | MODEL | BP |
Requests the Breusch-Pagan test for two-way random effects | MODEL | BP2 |
Requests the Bera, Sosa Escudero, and Yoon modified Rao’s score test | MODEL | BSY |
Specifies the between-groups model | MODEL | BTWNG |
<table>
<thead>
<tr>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<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 first-differences dynamic panel model</td>
<td>MODEL</td>
<td>DYNDIFF</td>
</tr>
<tr>
<td>Specifies the system dynamic panel model</td>
<td>MODEL</td>
<td>DYNSYS</td>
</tr>
<tr>
<td>Specifies the one-way fixed-effects model</td>
<td>MODEL</td>
<td>FIXONE</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>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 Hausman-Taylor model</td>
<td>MODEL</td>
<td>HTAYLOR</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 the 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 $\Phi$ matrix for the Parks method</td>
<td>MODEL</td>
<td>PHI</td>
</tr>
<tr>
<td>Specifies the pooled model</td>
<td>MODEL</td>
<td>POOLED</td>
</tr>
</tbody>
</table>
### Table 26.1 continued

<table>
<thead>
<tr>
<th>Description</th>
<th>Statement</th>
<th>Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>Requests poolability tests for one-way fixed effects and the 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 the 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 the panel unit root/stationarity test</td>
<td>MODEL</td>
<td>UROOTTEST=</td>
</tr>
<tr>
<td>Specifies the method for the variance components estimator</td>
<td>MODEL</td>
<td>VCOMP=</td>
</tr>
<tr>
<td>Specifies linear equality restrictions on the parameters</td>
<td>RESTRICT</td>
<td></td>
</tr>
<tr>
<td>Performs tests of linear hypotheses</td>
<td>TEST</td>
<td>WALD, LM, LR</td>
</tr>
<tr>
<td>Requests the Wooldridge (2002) test for the presence of unobserved effects</td>
<td>MODEL</td>
<td>WOOLDRIDGE02</td>
</tr>
</tbody>
</table>

**Comparing Models**

Create tables that display side-by-side model comparisons

**PROC PANEL Statement**

```sas
PROC PANEL options;
```

The PROC PANEL statement invokes the PANEL procedure. You can specify the following options:

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

- **OUTCOV**
  
  writes the standard errors and covariance matrix of the parameter estimates to the OUTEST= data set. For more information, see the section “OUTEST= Data Set” on page 1894.

- **OUTCORR**
  
  writes the correlation matrix of the parameter estimates to the OUTEST= data set. For more information, see the section “OUTEST= Data Set” on page 1894.
OUTEST=SAS-data-set
names an output data set to contain the parameter estimates. If you omit this option, the OUTEST=data set is not created. For more information about the structure of the OUTEST=data set, see the section “OUTEST= Data Set” on page 1894.

OUTTRANS=SAS-data-set
names an output data set to contain the transformed data. Several models that the PANEL procedure supports are estimated by first transforming the data and then applying standard regression techniques to the transformed data. This option enables you to access the transformed data. For more information about the structure of the OUTTRANS=data set, see the section “OUTTRANS= Data Set” on page 1895.

PLOTS < (global-plot-options < (NCROSS=value) > ) > < = (specific-plot-options) >
selects plots to be produced via the Output Delivery System. For general information about ODS Graphics, see Chapter 21, “Statistical Graphics Using ODS” (SAS/STAT User’s Guide). The global-plot-options apply to all relevant plots that the PANEL procedure generates.

Global Plot Options

The following global-plot-options are supported:

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

ONLY
suppresses the default plots. Only the plots that you specifically request are produced.

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

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.

PREDSSURFACE
produces a surface plot of predicted values.

QQ
produces a Q-Q plot of residuals.

RESIDSTACK | RESSTACK
produces a stacked plot of residuals.

RESIDSSURFACE
produces a surface plot of residual values.

RESIDUAL | RES
plots the residuals.

RESIDUALHISTOGRAM | RESIDHISTOGRAM
plots the histogram of residuals.
For more information, see the section “Creating ODS Graphics” on page 1893.

In addition, you can specify any of the following MODEL statement options in the PROC PANEL statement: AMACURDY, BTWNG, BTWNT, CORRB, COVB, DASILVA, DYNDIFF, DYN SYS, FDONE, FDone-Time, FDTWO, FIXONE, FIXONETIME, FIXTWO, HTAYLOR, M=, NOINT, NOPRINT, PARKS, PHI, POOLED, PRINTFIXED, RANONE, RANTWO, RHO, SINGULAR=, and VCOMP=. When you specify these options in the PROC PANEL statement, they apply globally to every MODEL statement. For a complete description of each of these options, see the section “MODEL Statement” on page 1818.

---

**BY Statement**

```
BY variables;
```

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

The following statements show an example:

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

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

---

**CLASS Statement**

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

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

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

---

**COMPARE Statement**

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

A COMPARE statement creates tables of side-by-side comparisons of parameter estimates and other model statistics. You can fit multiple models simultaneously by specifying multiple MODEL statements, and you can use a COMPARE statement to create tables that compare the models.
The COMPARE statement creates two tables: the first table compares model fit statistics such as R-square and mean square error; the second table compares regression coefficients, their standard errors, and (optionally) $t$ tests.

By default, comparison tables are created for all fitted models, but you can use the optional `model-list` to limit the comparison to a subset of the fitted models. The `model-list` consists of a set of model labels, as specified in the MODEL statement; for more information, see the section “MODEL Statement” on page 1818. If a model does not have a label, you refer to it generically as “Model $i$,” where the corresponding model is the $i$th MODEL statement specified. If model labels are longer than 16 characters, then only the first 16 characters of the labels in the `model-list` are used to determine a match.

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

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

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

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

**MSTAT(mstat-list)**

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

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

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

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

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

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

- **FNUMDF**
  displays the numerator degrees of freedom of the overall test for no fixed effects.
**COMPARE Statement**

- **M**
  displays the Hausman test \( m \) statistic.

- **MDF**
  displays the Hausman test degrees of freedom.

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

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

- **NONE**
  suppresses the table of model fit statistics.

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

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

- **PROBM**
  displays the significance level of the Hausman test.

- **RMSE**
  displays the model root mean square error.

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

- **SSE**
  displays the model sum of squares.

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

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

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

**OUTPARM=** *SAS-data-set*

names an output data set to contain the data from the comparison table for parameter estimates, standard errors, and \( t \) tests.

**OUTSTAT=** *SAS-data-set*

names an output data set to contain the data from the comparison table for model fit statistics, such as R-square and mean square error.
PSTAT(pstat-list)
specifies a list of parameter statistics to be displayed. By default, estimated regression coefficients and their standard errors are displayed. Use this option to specify a custom set of parameter statistics.

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

ALL
displays all parameter statistics.

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

NONE
suppresses the table of parameter statistics.

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

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

T
displays the $t$ statistic.

See Example 26.2 for a demonstration of the COMPARE statement.

FLATDATA Statement

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

The FLATDATA statement enables you to use PROC PANEL when you have data in flat (or wide) format, where all measurements for a particular cross section are contained within one observation. See Example 26.5 for a demonstration. If you have flat data, you should issue the FLATDATA statement first in PROC PANEL, before you reference any variables that you create using this statement.

You must specify the following options:

BASE=(basename basename . . . basename)
specifies the variables 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 only the base names, and the procedure extracts the appropriate variables to transform. If some year’s data are missing for a variable, then PROC PANEL detects this and fills in missing values.

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

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.

You can also specify the following option:
**KEEP=(variable variable ... variable)**

specifies the variables to be copied without any transformation. These variables remain constant with respect to time when the data are converted to PROC PANEL format.

You can also specify the following *option* after a slash (/):

**OUT=SAS-data-set**

saves the converted flat data set to a data set in PROC PANEL format.

---

### 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.

It is vitally important that you sort your data by cross sections and by time periods within cross sections. As PROC PANEL steps through the observations in the data, it treats any change in the value of the cross section ID variable as a new cross section, regardless of whether it has encountered that value previously. If you do not sort your data, the results might not be what you expect.

To make sure that the input data set is correctly sorted, use PROC SORT to sort the input data set, and use a BY statement to list the variables exactly as they are listed in the ID statement, as in the following example:

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

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

### INSTRUMENTS Statement

**INSTRUMENTS options ;**

The INSTRUMENTS statement is used in dynamic panel estimation (which you request via the DYNDIFF or DYNSYS option in the MODEL statement) to forgo the default set of instruments in favor of a custom set.

The INSTRUMENTS statement is also used to specify variables that are correlated with individual effects during Hausman-Taylor or Amemiya-MaCurdy estimation (which you request via the HTAYLOR or AMACURDY option, respectively, in the MODEL statement).

You can specify the following *options*:

**CONSTANT**

includes an intercept (column of ones) as an instrument in dynamic panel estimation.
**Chapter 26: The PANEL Procedure**

**CORRELATED=(variable variable ... variable)**
specifies a list of variables that are treated as correlated with the unobserved individual effects when you are fitting a Hausman-Taylor or Amemiya-MaCurdy model.

**DEPVAR<(DIFF | LEVEL | BOTH )>**
specifies instruments that are related to the dependent variable. You can specify the following values:

- **DIFF** creates instruments based on the dependent variable for the difference equations.
- **LEVEL** creates instruments based on the dependent variable for the level equations.
- **BOTH** creates instruments based on the dependent variable for the whole system.

The default is **BOTH**.

**DIFFEND=(variable variable ... variable)**
specifies a list of variables that are treated as endogenous when you are creating GMM-style instruments for the difference equations in dynamic panel estimation.

**DIFFEQ=(variable variable ... variable)**
specifies a list of variables that can be used as standard instruments for the difference equations in dynamic panel estimation.

**DIFFPRE=(variable variable ... variable)**
specifies a list of variables that are treated as predetermined when you are creating instruments for the difference equations in dynamic panel estimation.

**LEVELEND=(variable variable ... variable)**
specifies a list of variables that are treated as endogenous when you are creating instruments for the level equations in dynamic panel estimation.

**LEVELEQ=(variable variable ... variable)**
specifies a list of variables that can be used as standard instruments for the level equations in dynamic panel estimation.

**LEVELPRE=(variable variable ... variable)**
specifies a list of variables that are treated as predetermined when you are creating instruments for the level equations in dynamic panel estimation.

**MAXBAND=integer**
if specified, sets the maximum number of GMM-style instruments per observation, for each variable.

For a detailed discussion of the model setup and the use of the INSTRUMENTS statement for dynamic panel estimation, see the section “Dynamic Panel Estimation (DYNDIFF and DYNYSYS Options)” on page 1855.

For Hausman-Taylor or Amemiya-MaCurdy estimation, you specify which variables are correlated with the individual effects by using the CORRELATED= option. All other options are ignored. For these estimators, the specified variables are not instruments; they are merely designated as correlated. The instruments are determined by the method; for more information, see the section “Hausman-Taylor Estimation (HTAYLOR Option)” on page 1854.

When you specify multiple INSTRUMENT statements, each is paired with the MODEL statement that immediately follows. For example, the following statements fit two dynamic panel models that have custom instrumentation:
LAG, CLAG, SLAG, XLAG, and ZLAG Statements

LAG var\(_1\)\(\ldots\)var\(_N\)\((\text{lag}_1\ \text{lag}_2\ \ldots\ \text{lag}_T)\) / OUT=SAS-data-set;

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

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

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

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

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

ZLAG replaces missing values with 0 for that variable.

For all these alternative statements, missing values are replaced only if they are in the generated (lagged) series. Missing variables in the original variables remain unchanged.

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

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

If you want zeroing instead of missing values, then use the ZLAG statement in place of the LAG statement:

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

Similarly, you can use the XLAG statement to replace missing values with overall means, the SLAG statement to replace them with time means, and the CLAG statement to replace them with cross-sectional means.
**MODEL Statement**

```markdown
MODEL < "string"> response = regressors </options> ;
```

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

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

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

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

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

```markdown
MODEL "label" . . . ;
```

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

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

<table>
<thead>
<tr>
<th>Table 26.2</th>
<th>Summary of MODEL Statement Options</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Option</strong></td>
<td><strong>Description</strong></td>
</tr>
<tr>
<td><strong>Estimation Technique Options</strong></td>
<td></td>
</tr>
<tr>
<td>AMACURDY</td>
<td>Fits a one-way model by using the Amemiya-MaCurdy estimator</td>
</tr>
<tr>
<td>BTWNG</td>
<td>Fits the between-groups model</td>
</tr>
<tr>
<td>BTWNT</td>
<td>Fits the between-time-periods model</td>
</tr>
<tr>
<td>DASILVA</td>
<td>Fits a moving average model by using the Da Silva method</td>
</tr>
<tr>
<td>DYNDIFF</td>
<td>Fits a dynamic panel model by using GMM on the difference equations</td>
</tr>
<tr>
<td>DYNNSYS</td>
<td>Fits a dynamic panel model by using system GMM</td>
</tr>
<tr>
<td>FDONE</td>
<td>Fits a one-way model by using first-differenced methods</td>
</tr>
<tr>
<td>FDONETIME</td>
<td>Fits a one-way model for time effects by using first-differenced methods</td>
</tr>
<tr>
<td>FDTWO</td>
<td>Fits a two-way model by using first-differenced methods</td>
</tr>
<tr>
<td>FIXONE</td>
<td>Fits a one-way fixed-effects model</td>
</tr>
<tr>
<td>FIXONETIME</td>
<td>Fits a one-way fixed-effects model for time effects</td>
</tr>
<tr>
<td>FIXTWO</td>
<td>Fits a two-way fixed-effects model</td>
</tr>
</tbody>
</table>
Table 26.2  continued

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTAYLOR</td>
<td>Fits a one-way model by using the Hausman-Taylor estimator</td>
</tr>
<tr>
<td>PARKS</td>
<td>Fits an autoregressive model by using the Parks method</td>
</tr>
<tr>
<td>POOLED</td>
<td>Fits the pooled regression model</td>
</tr>
<tr>
<td>RANONE</td>
<td>Fits a one-way random-effects model</td>
</tr>
<tr>
<td>RANTWO</td>
<td>Fits a two-way random-effects model</td>
</tr>
</tbody>
</table>

**Estimation Control Options**

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>M=</td>
<td>Specifies the moving average order</td>
</tr>
<tr>
<td>NOESTIM</td>
<td>Limits estimation to only transforming the data</td>
</tr>
<tr>
<td>NOINT</td>
<td>Suppresses the intercept</td>
</tr>
<tr>
<td>SINGULAR=</td>
<td>Specifies a matrix inverse singularity criterion</td>
</tr>
<tr>
<td>VCOMP=</td>
<td>Specifies the type of variance component estimation for</td>
</tr>
<tr>
<td></td>
<td>random-effects estimation</td>
</tr>
</tbody>
</table>

**Dynamic Panel Estimation Control Options**

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARTEST=</td>
<td>Specifies the maximum order of the autoregression (AR) test</td>
</tr>
<tr>
<td>ATOL=</td>
<td>Specifies the convergence criterion of iterated GMM, with respect</td>
</tr>
<tr>
<td></td>
<td>to the weighting matrix</td>
</tr>
<tr>
<td>BIASCORRECTED</td>
<td>Requests bias-corrected variances for two-step GMM</td>
</tr>
<tr>
<td>BTOL=</td>
<td>Specifies the convergence criterion of iterated GMM, with respect</td>
</tr>
<tr>
<td></td>
<td>to the parameter matrix</td>
</tr>
<tr>
<td>DLAGS=</td>
<td>Specifies the number of dependent variables to be used as</td>
</tr>
<tr>
<td></td>
<td>regressors</td>
</tr>
<tr>
<td>GINV=</td>
<td>Specifies the type of generalized matrix inverse</td>
</tr>
<tr>
<td>GMM1</td>
<td>Estimates by one-step GMM, the default</td>
</tr>
<tr>
<td>GMM2</td>
<td>Estimates by two-step GMM</td>
</tr>
<tr>
<td>ITGMM</td>
<td>Estimates by iterative GMM</td>
</tr>
<tr>
<td>MAXITER=</td>
<td>Specifies the maximum iterations for iterative GMM</td>
</tr>
<tr>
<td>ROBUST</td>
<td>Specifies the robust covariance matrix</td>
</tr>
<tr>
<td>TIME</td>
<td>Includes time dummy variables in the model</td>
</tr>
</tbody>
</table>

**Alternative Variances Options**

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLUSTER</td>
<td>Corrects covariance for intracluster correlation</td>
</tr>
<tr>
<td>HAC(options)</td>
<td>Specifies a heteroscedasticity- and autocorrelation-consistent</td>
</tr>
<tr>
<td></td>
<td>(HAC) covariance</td>
</tr>
<tr>
<td>HCCME=</td>
<td>Specifies a heteroscedasticity-corrected covariance matrix</td>
</tr>
<tr>
<td></td>
<td>estimator (HCCME)</td>
</tr>
<tr>
<td>Option</td>
<td>Description</td>
</tr>
<tr>
<td>------------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>NEWEYWEST(\textit{options})</td>
<td>Specifies the Newey-West covariance, a special case of the HAC covariance</td>
</tr>
</tbody>
</table>

**Unit Root Test Options**

- \texttt{UROOTTEST(test-options)} Requests one or more panel data unit root and stationarity tests; specify \textit{test-options} ALL through ILC within this option.
- \texttt{STATIONARITY(test-options)} Synonym for the UROOTTEST option
- \texttt{ALL} Requests that all unit root tests be performed
- \texttt{BREITUNG(options)} Specifies Breitung’s tests that are robust to cross-sectional dependence
- \texttt{COMBINATION(options)} Specifies one or more unit root tests that combine over all cross sections
- \texttt{FISHER(options)} Synonym for the COMBINATION option
- \texttt{HADRI(options)} Specifies Hadri’s (2000) stationarity test
- \texttt{HT} Specifies the Harris and Tzavalis (1999) panel unit root test
- \texttt{IPS(options)} Specifies the Im, Pesaran, and Shin (2003) panel unit root test
- \texttt{LLC(options)} Specifies the Levin, Lin, and Chu (2002) panel unit root test

**Model Specification Test Options**

- \texttt{BFN} Requests the $R_\rho$ statistic for serial correlation under fixed effects
- \texttt{BL91} Requests the Baltagi and Li (1991) Lagrange multiplier (LM) test for serial correlation and random effects
- \texttt{BL95} Requests the Baltagi and Li (1995) LM test for first-order correlation under fixed effects
- \texttt{BP} Requests the Breusch-Pagan one-way test for random effects
- \texttt{BP2} Requests the Breusch-Pagan two-way test for random effects
- \texttt{BSY} Requests the Bera, Sosa Escudero, and Yoon modified Rao’s score test
- \texttt{BW} Requests the Berenblut-Webb statistic for serial correlation under fixed effects
- \texttt{CDTEST(options)} Requests a battery of cross-sectional dependence tests
- \texttt{DW} Requests the Durbin-Watson statistic for serial correlation under fixed effects
- \texttt{GHM} Requests the Gourieroux, Holly, and Monfort test for two-way random effects
- \texttt{HONDA} Requests the Honda one-way test for random effects
- \texttt{HONDA2} Requests the Honda two-way test for random effects
### Table 26.2  *continued*

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>KW</td>
<td>Requests the King and Wu two-way test for random effects</td>
</tr>
<tr>
<td>POOLTEST</td>
<td>Requests poolability tests for one-way fixed effects and pooled models</td>
</tr>
<tr>
<td>WOOLDRIDGE02</td>
<td>Requests the Wooldridge (2002) test for unobserved effects</td>
</tr>
</tbody>
</table>

#### Printed Output Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>CORR</td>
<td>Prints the parameter correlation matrix</td>
</tr>
<tr>
<td>CORRB</td>
<td>Synonym for the CORR option</td>
</tr>
<tr>
<td>COVB</td>
<td>Prints the parameter covariance matrix</td>
</tr>
<tr>
<td>ITPRINT</td>
<td>Prints the iteration history</td>
</tr>
<tr>
<td>NOPRINT</td>
<td>Suppresses normally printed output</td>
</tr>
<tr>
<td>PHI</td>
<td>Prints the $\Phi$ covariance matrix for the Parks method</td>
</tr>
<tr>
<td>PRINTFIXED</td>
<td>Estimates and prints the fixed effects</td>
</tr>
<tr>
<td>RHO</td>
<td>Prints the autocorrelation coefficients for the Parks method</td>
</tr>
<tr>
<td>VAR</td>
<td>Synonym for the COVB option</td>
</tr>
</tbody>
</table>

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

#### Estimation Technique Options

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

All estimation methods are described in the section “Details: PANEL Procedure” and its subsections.

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

**BTWNG**
estimates a between-groups model.

**BTWNT**
estimates a between-time-periods model.

**DASILVA**
estimates the model by using the Da Silva method, which assumes a mixed variance-component moving average model for the error structure.
DYNDIFF
estimates a dynamic panel model by the generalized method of moments (GMM), performed on the
difference equations. A default set of instruments is assumed. You can optionally specify your own
instruments by using an INSTRUMENTS statement.

DYNZY
estimates a dynamic panel model by the generalized method of moments (GMM), performed on the
system of both the differenced and level equations. A default set of instruments is assumed. You can
optionally specify your own instruments by using an INSTRUMENTS statement.

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

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

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

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

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

FIXTWO
estimates a two-way fixed-effects model.

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

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

POOLED
estimates a pooled (OLS) model.

RANONE
estimates a one-way random-effects model.

RANTWO
estimates a two-way random-effects model.

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

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

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

- **NOINT**
  suppresses the intercept parameter from the model.

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

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

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

  By default, VCOMP=FB for balanced data and VCOMP=WK for unbalanced data. For more information, see the sections “One-Way Random-Effects Model (RANONE Option)” on page 1843 and “Two-Way Random-Effects Model (RANTWO Option)” on page 1846.

**Dynamic Panel Estimation Control Options**

Dynamic panel estimation control options are specific to dynamic panels, where the estimation technique is specified as DYNDIFF or DYNSYS. For more information, see the section “Dynamic Panel Estimation (DYNDIFF and DYNSYS Options)” on page 1855.

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

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

- **BIASCORRECTED**
  computes the bias-corrected covariance matrix of the two-step dynamic panel estimator. When you specify this option, the ROBUST option is disabled for the two-step GMM estimator.
**BTOL=number**
specifies the convergence criterion for iterated GMM when convergence of the method is determined by convergence in the parameter matrix. The convergence criterion (number) must be positive. By default, BTOL=1E–8.

**DLAGS=number**
specifies the number of dependent-variable lags to use as regressors. By default, DLAGS=1.

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

- **G2** uses the G2 generalized inverse.
- **G4** uses the G4 generalized inverse.

The difference between G2 and G4 becomes evident when you invert singular matrices. The G2 generalized inverse drops rows and columns from singular matrices to produce a viable inverse. The G4 inverse, on the other hand, is the Moore-Penrose generalized inverse. The Moore-Penrose generalized inverse averages the variance effects between collinear rows. The G4 inverse is usually more stable, but it is computationally intensive. By default, GINV=G2. If you have trouble reproducing published results, often the solution is to switch to GINV=G4.

**GMM1**
estimates the dynamic panel regression by the one-step generalized method of moments (GMM). This is the default estimation method.

**GMM2**
estimates the dynamic panel regression by two-step GMM.

**ITGMM**
estimates the dynamic panel regression by iterative GMM.

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

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

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

### Alternative Variances Options

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

**CLUSTER**
specifies the cluster correction for the covariance matrix. You can specify this option when you specify HCCME=0, 1, 2, or 3.
HAC < (options) >
specifies the heteroscedasticity- and autocorrelation-consistent (HAC) covariance matrix estimator. This option is not available for between models and cannot be combined with the HCCME= option.

For more information, see the section “Heteroscedasticity- and Autocorrelation-Consistent Covariance Matrices” on page 1870.

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

- **ADJUSTDF**
  makes a small-sample adjustment to the degrees of freedom in the covariance calculation.

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

  number
  specifies a fixed value of the bandwidth parameter.

  **ANDREWS91 | ANDREWS**

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

- **SAMPLESIZE<(options)>**
  **SS<(options)>**
  calculates the bandwidth according to the following equation based on the sample size,

  \[ b = \gamma T^r + c \]

  where \( b \) is the bandwidth parameter; \( T \) is the sample size; and \( \gamma \), \( r \), and \( c \) are values that you specify using the following options within parentheses and separated by commas:

  **CONSTANT=number**
  specifies the constant \( c \) in the equation. By default, CONSTANT=0.5.

  **GAMMA=number**
  specifies the coefficient \( \gamma \) in the equation. By default, GAMMA=0.75.

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

  **RATE=number**
  specifies the growth rate \( r \) in the equation. By default, RATE=0.3333.

By default, BANDWIDTH=ANDREWS91.
KERNEL=BARTLETT | PARZEN | QS | TH | TRUNCATED

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

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

By default, KERNEL=TRUNCATED.

KERNELLB=number

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

PREWHITENING

requests prewhitening in the covariance calculation.

HCCME=NO | number

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

NO does not correct the covariance matrix.
number specifies the type of covariance adjustment. The value of number can be any integer from 0 to 4, inclusive.

For more information, see the section “Heteroscedasticity-Corrected Covariance Matrices” on page 1866. By default, HCCME=NO.

NEWEYWEST<(options)>

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

CONSTANT=number

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

GAMMA=number

specifies the coefficient \( \gamma \) in the equation. By default, GAMMA=0.75.

RATE=number

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

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

\[ \text{NEWEYWEST} \]
PROC PANEL DATA=A;
  ID i t;
  MODEL y = x1 x2 x3 / ranone hac(kernel = bartlett bandwidth = 3);
  MODEL y = x1 x2 x3 / ranone neweywest(gamma = 0, constant = 3);
RUN;

Unit Root Test Options

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

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

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

You can specify the following tests and test-options:

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

DETAIL
prints intermediate results (lag order).

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

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

AIC
selects the order of lags by Akaike’s information criterion (AIC).
selects the order of lags by Hall’s (1994) sequential testing method, beginning with the most general model (maximum lags) and then reducing lag orders sequentially.

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

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

SBC

SIC

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

SG
selects the order of lags by Hall’s (1994) sequential testing method, beginning with no lag terms and then increasing lag orders sequentially.

By default, LAG=MAIC.

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

COMBINATION < (test-options) >
FISHER < (test-options) >
specifies combination tests that are proposed by Choi (2001); Maddala and Wu (1999). You can specify one or more of the following test-options within parentheses and separated by spaces:

TEST=ADF | PP
selects the time series unit root test for combination tests. You can specify the following values:

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

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

By default, TEST=PP.

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

BARTLETT
specifies the Bartlett kernel function.

PARZEN
specifies the Parzen kernel function.
QS specifies the quadratic spectral kernel function.

TH specifies the Tukey-Hanning kernel function.

TRUNCATED specifies the truncated kernel function.

By default, KERNEL=QS.

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

ANDREWS selects the bandwidth by the Andrews method.

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

By default, BANDWIDTH=ANDREWS.

DETAIL
prints intermediate results (lag order and long-run variance for each cross section).

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

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

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

GS selects the order of lags by Hall's (1994) sequential testing method, beginning with the most general model (maximum lags) and then reducing lag orders sequentially.

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

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

SBC

SIC

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

SG selects the order of lags by Hall's (1994) sequential testing method, beginning with no lag terms and then increasing lag orders sequentially.

By default, LAG=MAIC.
MAXLAG=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 LAG=value.

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

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

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

By default, KERNEL=QS.

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

  ANDREWS        selects the bandwidth by the Andrews method.
  number         sets the bandwidth to number, which must be nonnegative.

By default, BANDWIDTH=ANDREWS.

DETAIL
prints intermediate results (lag order and long-run variance for each cross section).

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

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

DETAIL
prints intermediate results (lag order).

LAG=type | value
specifies the method to choose the lag order for the augmented Dickey-Fuller (ADF) regres-
sions. You can specify a value or one of the types listed after value.
value
specifies the lag order. If the lag order is too big to run linear regression \((\text{value} > T - k)\),
where \(T\) is the number of time periods and \(k\) is the number of parameters), then the lag
order is set to \(12(T/100)^{1/4}\) or \(T - k - 1\), whichever is smaller.

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

GS
selects the order of lags by Hall’s (1994) sequential testing method, beginning with the
most general model (maximum lags) and then reducing lag orders sequentially.

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

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

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

SG
selects the order of lags by Hall’s (1994) sequential testing method, beginning with no
lag terms and then increasing lag orders sequentially.

By default, LAG=MAIC.

MAXLAG=value
specifies the maximum lag order that the model allows. The default value is \(12(T/100)^{1/4}\).
If \(\text{value}\) 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 LAG=\(\text{value}\).

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

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

BARTLETT
specifies the Bartlett kernel function.

PARZEN
specifies the Parzen kernel function.

QS
specifies the quadratic spectral kernel function.

TH
specifies the Tukey-Hanning kernel function.

TRUNCATED
specifies the truncated kernel function.

By default, KERNEL=QS.
Chapter 26: The PANEL Procedure

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

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

DETAIL
prints intermediate results (lag order and long-run variance for each cross section).

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

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

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

GS selects the order of lags by Hall’s (1994) sequential testing method, beginning with the most general model (maximum lags) and then reducing lag orders sequentially.

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

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

SBC SIC SBIC

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

SG selects the order of lags by Hall’s (1994) sequential testing method, beginning with no lag terms and then increasing lag orders sequentially.

By default, LAG=MAIC.

MAXLAG=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 LAG=value.

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

\[
\text{BANDWIDTH=ANDREWS | number}
\]

\[
\text{DETAIL}
\]

\[
\text{LAG=type | value}
\]

\[
\text{MAXLAG=value}
\]
proc panel data=A;
  id i t;
  model y = x1 x2 x3 / unitroot(llc(kernel = parzen lag = aic),
    breitung(lag = gs)
    maxlag = 2
    kernel = bartlett);
run;

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

Model Specification Test Options

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

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

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

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

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

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

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

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

**CDTEST < (P=value) >**
requests cross-sectional dependence tests. These include the Breusch and Pagan (1980) LM test, the scaled version of the Breusch and Pagan (1980) test, and the Pesaran (2004) CD test. When you specify P=value, the CD test for local cross-sectional dependence is performed using the order value, where value is an integer greater than 0.
DW (Experimental) requests the Durbin-Watson statistic for serial correlation under cross-sectional fixed effects.

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

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

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

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

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

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

Printed Output Options

Printed output options change how results are presented.

**CORRB**

**CORR**

prints the matrix of estimated correlations between the parameter estimates.

**COVB**

**VAR**

prints the matrix of estimated covariances between the parameter estimates.

**ITPRINT**

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

**NOPRINT**

suppresses the normal printed output.

**PHI**

prints the $\Phi$ matrix of estimated covariances of the observations for the Parks method. The PHI option is relevant only when you specify the PARKS option. For more information, see the section “Parks Method for Autoregressive Models (PARKS Option)” on page 1849.

**PRINTFIXED**

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

**RHO**

prints the estimated autocorrelation coefficients for the Parks method.
OUTPUT Statement

\[ \text{OUTPUT} \ < \text{options}> \ ; \]

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

\[ \text{OUT=} \text{SAS-data-set} \]

names the output SAS data set to contain the predicted and transformed values. If you do not specify this option, the new data set is named according to the DATA\& convention.

\[ \text{PREDICTED=} \text{name} \]

\[ \text{P=} \text{name} \]

writes the predicted values to the output data set.

\[ \text{RESIDUAL=} \text{name} \]

\[ \text{R=} \text{name} \]

writes the residuals to the output data set.

RESTRICT Statement

\[ \text{RESTRICT} \ < \text{"string"} > \text{ equation} < , \text{equation2} . . . > ; \]

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

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

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

The following statements illustrate the use of the RESTRICT statement:

```sas
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;
```

If you are fitting a dynamic panel model, you can place restrictions on lags of the dependent variable by referencing the name of the dependent variable followed by an underscore and the lag order. For example,
proc panel;
    model sales = price / dyndiff;
    restrict sales_1 = 0.5;
run;

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

TEST Statement

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

The TEST statement performs Wald, Lagrange multiplier, and likelihood ratio tests of linear hypotheses about the regression parameters in the preceding MODEL statement. Like RESTRICT statements, TEST statements before the first MODEL statement are automatically associated with the first MODEL statement, as are any TEST statements that follow it but precede subsequent MODEL statements. Each equation specifies a linear hypothesis to be tested. All hypotheses in one TEST statement are tested jointly. Variable names in the equations must correspond to regressors in the preceding MODEL statement, and each name represents the coefficient of the corresponding regressor. The keyword INTERCEPT refers to the coefficient of the intercept.

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

ALL
    specifies Wald, Lagrange multiplier, and likelihood ratio tests.

LM
    specifies the Lagrange multiplier test.

LR
    specifies the likelihood ratio test.

WALD
    specifies the Wald test.

The Wald test is performed by default.

The following statements illustrate the use of the TEST statement:

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

The first test investigates the joint hypothesis that

\[ \beta_1 = 0 \]

and

\[ 0.5\beta_2 + 2\beta_3 = 0 \]
Currently, only linear equality restrictions and tests are permitted in PROC PANEL. Tests and restriction expressions can be composed only of algebraic operations that involve the addition symbol (+), subtraction symbol (–), and multiplication symbol (*).

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

If you are fitting a dynamic panel model, you can perform tests on lags of the dependent variable by referencing the name of the dependent variable followed by an underscore and the lag order. For example,

```plaintext
proc panel;
    model sales = price / dyndiff;
    test sales_1 = 0.5 / wald;
run;
```

For the Da Silva, Hausman-Taylor, Amemiya-MacCurdy, and dynamic panel methods, only the Wald test is available.

### Details: PANEL Procedure

#### Specifying the Input Data

Panel data are identified by both a cross section identification (ID) variable and a time variable. Suppose that you have a data set Sample, where cross sections are identified by the variable State and time periods are identified by the variable Date. The input data set that PROC PANEL uses must be sorted by cross section and by time within each cross section. As PROC PANEL steps through the observations in the data, it treats any change in the value of the cross section ID variable as a new cross section, regardless of whether it has encountered that value previously. If you do not sort your data, the results might not be what you expect. 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 Sample appropriately:

```plaintext
proc sort data=sample;
    by state date;
run;
```

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

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

Alternatively, PROC PANEL has the capability to read flat (or wide) data. Suppose you are using the data set Flat, which has observations on states. Specifically, the data are composed of observations on Y, X1, and X2. Unlike the data in the Sample data set, these data are not long. Instead, you have all of a state’s information
in a single row. The time observations for the \textit{Y} variable are recorded horizontally. So the variable \textit{Y\_1} is the first period’s time observation, and the variable \textit{Y\_10} is the tenth period’s observation for some state. The same is true of the other variables. You have the variables \textit{X1\_1} through \textit{X1\_10} and \textit{X2\_1} through \textit{X2\_10}. For such data, use 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;
```

For more information about the FLATDATA statement, see the section “FLATDATA Statement” on page 1814 and Example 26.5.

---

### Specifying the Regression Model

The PANEL procedure is similar to other regression procedures in SAS software. Suppose you want to regress the variable \textit{Y} on the regressors \textit{X1} and \textit{X2}. You specify the dependent variable first, followed by an equal sign, followed by the list of regression variables, as shown in the following statements:

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

One advantage of using PROC PANEL is that you can incorporate a model for the structure of the error terms. It is important to consider what type of model is appropriate for your data and to specify the corresponding option in the MODEL statement. The following model estimation options are supported: POOLED, BTWNG, BTWNT, FIXONE, FIXONETIME, FIXTWO, FDONE, FDONETIME, FDTWO, RANONE, RANTWO, PARKS, DASILVA, HTAYLOR, AMACURDY, DYNDIFF, and DYNYSYS. The methods that underlie these estimation options are described in this same order, beginning with the section “Pooled Regression (POOLED Option)” on page 1840.

The following statements fit a one-way random-effects model with variance components estimated by the Fuller-Battese (FB) method:

```sas
proc panel data=sample;
   id State Date;
   model Y = X1 X2 / ranone vcomp = fb;
run;
```

You can specify more than one estimation option in the MODEL statement, and the analysis is repeated for each specified method. You can use multiple MODEL statements to estimate different regression models or to estimate the same model by different methods.

The DYNDIFF and DYNYSYS options cannot be combined with other estimation options in the MODEL statement. If you want to fit a dynamic panel model and perform some other estimation (such as one-way random effects), specify multiple MODEL statements.
Missing Values

Any observation in the input data set that has a missing value for the cross section ID, time series ID, dependent variable, or any model effect is ignored by PROC PANEL when it fits the model.

If your data contain observations in which only the dependent variable is missing, you can still compute predicted values for these observations and store them in an output data set by using the OUTPUT statement.

Unbalanced Data

Unbalanced data occur when not all time values are observed for all cross sections or, if time is not part of the estimation, when the cross sections are not all the same size.

Whether the data are unbalanced by design or because of missing values, almost all the methods that the PANEL procedure supports take proper account of the unbalanced data. The lone exceptions are the Amemiya-MaCurdy, Da Silva, and Parks methods, which are suitable only for balanced data.

Common Notation

This section presents notation that is common to all subsequent sections. Consider the panel regression:

\[ y_{it} = \alpha + \sum_{k=1}^{K} x_{ikt} \beta_k + u_{it} \quad i = 1, \ldots, N; \quad t = 1, \ldots, T_i \]

The total number of observations is \( M = \sum_{i=1}^{N} T_i \). For balanced data, \( T_i = T \) for all \( i \). For unbalanced data, define \( T \) to be the number of unique time periods.

The exact representation of \( u_{it} \) and the underlying assumptions depend on the estimation method.

In matrix notation the model is

\[ y_{it} = \alpha + x_{it} \beta + u_{it} \]

where \( x_{it} \) is a \( 1 \times K \) row vector of independent variables and \( \beta \) is the \( K \times 1 \) vector of coefficients. Let \( y \) and \( X \) be matrices that are formed by arranging the dependent and independent variables by cross section, and by time within each cross section. Let \( X_{\alpha} \) be the \( X \) matrix augmented by a first column of ones, which corresponds to the intercept term \( \alpha \).

Define the following utility matrices:

- \( I_p \) is an identity matrix of dimension \( p \).
- \( j_p \) is a \( p \times 1 \) column vector of ones.
- \( J_p = j_p j_p' \) is a matrix of ones of dimension \( p \).
- \( \bar{J}_p = p^{-1} J_p \).
\[ E_p = I_p - \hat{J}_p. \]

In the following sections, the panel data are assumed to be unbalanced unless otherwise indicated. If the data are balanced, the formulas reduce appropriately.

### Pooled Regression (POOLED Option)

You perform pooled regression by specifying the POOLED option in the MODEL statement. Pooled regression is standard ordinary least squares (OLS) regression without any cross-sectional or time effects. The error structure is simply \( u_{it} = e_{it} \), where the \( e_{it} \) are independently and identically distributed (iid) with zero mean and variance \( \sigma_e^2 \).

### Between-Groups Regression (BTWNG and BTWNT Options)

You perform between-groups regression by specifying the BTWNG option in the MODEL statement. Between-groups regression is ordinary least squares (OLS) regression performed on data that have been collapsed into cross-sectional means.

The BTWNT option works similarly, except that the data are collapsed by time period instead of by cross section.

### One-Way Fixed-Effects Model (FIXONE and FIXONETIME Options)

You perform one-way fixed-effects estimation by specifying the FIXONE option in the MODEL statement. The error structure for the one-way fixed-effects model is

\[ u_{it} = v_i + e_{it} \]

where the \( v_i \) are nonrandom parameters that are restricted to sum to 0, and the \( e_{it} \) are iid with zero mean and variance \( \sigma_e^2 \).

The fixed-effects model can be estimated by ordinary least squares (OLS), treating the \( v_i \) as coefficients on dummy variables that identify the cross sections. However, when \( N \) is large, you might want to estimate only \( \beta \) and not \( v_i \).

Let \( Q_0 = \text{diag}(E_{T-1}) \). The matrix \( Q_0 \) represents the within transformation, the conversion of the raw data to deviations from a cross section’s mean. Let \( X_w = Q_0 X \) and \( y_w = Q_0 y \). The within estimator of \( \beta \) is

\[ \hat{\beta}_w = (X_{w}^\prime X_{w})^{-1} X_{w}^\prime y_w \]

The previous estimation does not involve the intercept term because \( \hat{\beta}_w \) is the same whether or not the intercept \( \alpha \) is included in the model.

Standard errors, \( t \) statistics, and fit statistics such as mean square error (MSE) are all equivalent to those obtained from OLS regression of \( y_w \) on \( X_w \). The only exception is the error degrees of freedom, which equals \( M - N - K \) to account for the tacit estimation of the \( N \) fixed effects.
Each fixed effect is estimated as

$$\hat{v}_t = \bar{y}_t - \bar{X}_t \hat{\beta}_w$$

where $\bar{y}_t$ and $\bar{X}_t$ are cross-sectional means.

The fixed-effects model is parameterized so that the intercept is the fixed effect for the last cross section. That is,

$$\hat{\alpha} = \hat{\nu}_N = \bar{y}_N - \bar{X}_N \hat{\beta}_w$$

Fixed effects are by default not displayed as part of the regression, but you can obtain them by specifying the PRINTFIXED option in the MODEL statement. In models that have an intercept, the printed fixed effects are the deviations $\nu_t - \nu_N$. To display the untransformed fixed effects, specify both the NOINT and PRINTFIXED options.

Variance estimates of $\hat{\alpha}$, $\hat{\nu}_t$, and $\nu_t - \nu_N$ are obtained by the delta method.

The FIXONETIME option works similarly, except that the data are grouped by time period instead of by cross section.

---

**Two-Way Fixed-Effects Model (FIXTWO Option)**

You perform two-way fixed-effects estimation by specifying the FIXTWO option in the MODEL statement. The error specification for the two-way fixed-effects model is

$$u_{it} = v_i + \lambda_t + e_{it}$$

where the $v_i$ and $\lambda_t$ are nonrandom parameters to be estimated.

Estimation is similar to that for one-way fixed effects, for which a within transformation is used to convert the problem to OLS regression. For two-way models under the general case of unbalanced data, the within transformation is more complex.

Following Wansbeek and Kapteyn (1989) and Baltagi (2013, sec. 9.4), let $X^*$ and $y^*$ be versions of $X$ and $y$ whose rows are sorted by time period, and by cross section within each time period. With the data sorted in this manner, define $D_N$ to be the $M \times N$ design matrix for cross sections. Each row of $D_N$ contains a 1 in the column that corresponds to that observation’s cross section, and 0s in the remaining columns. Similarly, define $D_T$ to be the $M \times T$ design matrix for time periods. In balanced data, $D_N = j_T \otimes I_N$ and $D_T = I_T \otimes j_N$.

Define the following:

$$\Delta_N = D_N' D_N \quad (N \times N)$$

$$\Delta_T = D_T' D_T \quad (T \times T)$$

$$A = D_T' D_N \quad (T \times N)$$

$$\tilde{D} = D_T - D_N \Delta_N^{-1} A' \quad (M \times N)$$

$$Q = \Delta_T - A \Delta_N^{-1} A' \quad (T \times T)$$

$$P = I_M - D_N \Delta_N^{-1} D_N' - \tilde{D} Q^{-1} \tilde{D}' \quad (M \times M)$$
The matrix \( P \) provides the two-way within transformation. If the data are balanced, this amounts to transforming any data value \( z_{it} \) to \( z_{it} - \bar{z}_{i} - \bar{z}_{t} + \bar{z}_{..} \).

Applying the two-way within transformation means that you can use OLS regression of \( Py^{*} \) on \( PX^{*} \) to obtain \( \tilde{\beta}_f \), \( \text{Var}(\tilde{\beta}_f) \), and fit statistics such as mean square error (MSE), provided that you adjust the error degrees of freedom to equal \( M - N - T - K + 1 \).

Define the residual vector \( r^{*} = y^{*} - X^{*} \tilde{\beta}_f \). Estimates of the time effects are \( \hat{\lambda} = Q^{-1} \tilde{D} r^{*} \), and estimates of the cross-sectional effects are \( \hat{v} = (\Theta_1 - \Theta_2 + \Theta_3) r^{*} \), where

\[
\begin{align*}
\Theta_1 &= \Delta_N^{-1} D_N' \\
\Theta_2 &= \Delta_N^{-1} A' Q^{-1} D_T' \\
\Theta_3 &= \Delta_N^{-1} A' Q^{-1} A \Delta_N^{-1} D_N'
\end{align*}
\]

The full model that contains the intercept, \( N \) cross-sectional effects, and \( T \) time effects is overidentified, and simultaneous estimation of these quantities is not possible without restrictions. If you specify the PRINTFIXED option, the printed fixed effects reflect these restrictions.

If the model has an intercept, then the PRINTFIXED option output is parameterized as follows:

- Intercept: \( \hat{v}_N + \hat{\lambda}_T \)
- Cross section \( i \): \( \hat{v}_i - \hat{v}_N \)
- Time period \( t \): \( \hat{\lambda}_t - \hat{\lambda}_T \)

If the model does not include an intercept, then the PRINTFIXED option output is parameterized as follows:

- Cross section \( i \): \( \hat{v}_i + \hat{\lambda}_T \)
- Time period \( t \): \( \hat{\lambda}_t - \hat{\lambda}_T \)

Variance and covariance estimates for the intercept and printed fixed effects are obtained by the delta method, because each of these quantities is a linear transformation of \( y^{*} \) and \( \tilde{\beta}_f \).

### One-Way Fixed-Effects Model, First Differencing (FDONE and FDONETIME Options)

You perform one-way fixed-effects estimation via first differencing by specifying the FDONE option in the MODEL statement. The method of first differencing offers an alternative to the within estimator \( \tilde{\beta}_w \).

Consider the following one-way fixed-effects model:

\[
y_{it} = \alpha + x_{it} \beta + v_i + e_{it}
\]

For this model, the fixed effects are removed by subtracting first-order lags from both sides of the equation:

\[
y_{it} - y_{i, t-1} = (x_{it} - x_{i, t-1}) \beta + (e_{it} - e_{i, t-1})
\]
Define $\Delta y_{it} = y_{i,t} - y_{i,t-1}$ and $\Delta x_{it} = x_{i,t} - x_{i,t-1}$, for $i = 1, \ldots, N$ and $t = 2, \ldots, T_i$. You obtain the first-differenced estimator, $\hat{\beta}_d$, and its variance by performing OLS regression of $\Delta y_{it}$ on $\Delta x_{it}$.

The estimation and parameterization of $(\alpha, v_i)$ are identical to that described in the section “One-Way Fixed-Effects Model (FIXONE and FIXONETIME Options)” on page 1840, with $\hat{\beta}_w$ replaced by $\hat{\beta}_d$.

The FDONETIME option works similarly, switching the roles of cross sections and time periods in the methodology described previously.

---

**Two-Way Fixed-Effects Model, First Differencing (FDTWO Option)**

You perform two-way fixed-effects estimation via first differencing by specifying the FDTWO option in the MODEL statement. The method of first differencing offers an alternative to the within estimator $\hat{\beta}_f$. Consider the following two-way fixed-effects model:

$$y_{it} = \alpha + x_{it} \beta + v_i + \lambda_t + e_{it}$$

For this model, the fixed effects are removed by the transformations $\Delta y_{it} = y_{i,t} - y_{i,t-1}$, $t - y_{i,t-1} = y_{i-1,t-1}$, and $\Delta x_{it} = x_{i,t} - x_{i,t-1}$, $t - x_{i,t-1} = x_{i-1,t-1}$. You obtain the two-way first-differenced estimator, $\hat{\beta}_{fd}$, and its variance by performing OLS regression of $\Delta y_{it}$ on $\Delta x_{it}$.

The estimation and parameterization of $(\alpha, v_i, \lambda_t)$ are identical to that described in the section “Two-Way Fixed-Effects Model (FIXTWO Option)” on page 1841, with $\hat{\beta}_f$ replaced by $\hat{\beta}_{fd}$.

---

**One-Way Random-Effects Model (RANONE Option)**

You perform one-way random-effects estimation by specifying the RANONE option in the MODEL statement. The specification for the one-way random-effects model is

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

where the $v_i$ are iid with zero mean and variance $\sigma^2_v$, and the $e_{it}$ are iid with zero mean and variance $\sigma^2_e$. Furthermore, a random-effects specification assumes that the error terms are mutually uncorrelated and that each error term is uncorrelated with $X$.

Estimation proceeds in two steps. First, you obtain estimates of the variance components $\sigma^2_v$ and $\sigma^2_e$. Second, with the variance components in hand, you form a weight for each cross section,

$$\hat{\theta}_i = 1 - \hat{\sigma}_e/\hat{\sigma}_v$$

where $\hat{\sigma}_v^2 = T_i \hat{\sigma}_v^2 + \hat{\sigma}_e^2$. Taking $\hat{\theta}_i$, you form the partial deviations:

$$\tilde{y}_{it} = y_{it} - \hat{\theta}_i \bar{y}_i.$$  
$$\tilde{x}_{\alpha, it} = x_{\alpha, it} - \hat{\theta}_i \bar{x}_{\alpha, i}.$$  

The random-effects estimation is then the result of OLS regression on the transformed data.

The PANEL procedure provides four methods of estimating variance components, as described in the following subsections.
Wallace-Hussain Method

You can use the Wallace-Hussain (1969) method of estimating variance components by specifying the VCOMP=WH option in the MODEL statement. The Wallace-Hussain method is part of a class of methods known as analysis of variance (ANOVA) estimators.

ANOVA estimators obtain variance components by solving a system of equations that is based on expected sums of squares. The following quadratic forms correspond to the within and between sums of squares, respectively:

\[ q_e = u'Q_0u \]
\[ q_v = u'P_0u \]

In these equations, \( Q_0 = \text{diag}(E_T)_i \), \( P_0 = \text{diag}(J_T)_i \), and \( u \) is the vector of true residuals.

The ANOVA methods differ only in how they estimate \( u \). The Wallace-Hussain method uses the residuals from pooled (OLS) regression, \( \hat{u}_p \), in both quadratic forms.

The expected values of the quadratic forms are

\[ E(\hat{u}_p'Q_0\hat{u}_p) = (d_1 - d_3)\sigma_v^2 + (M - N - K - 1 + d_2)\sigma_e^2 \]
\[ E(\hat{u}_p'P_0\hat{u}_p) = (M - 2d_1 + d_3)\sigma_v^2 + (N - d_2)\sigma_e^2 \]

where

\[ d_1 = \text{tr}\left\{ (X'\alpha X_\alpha)^{-1}X'_\alpha Z_0 Z'_0 X_\alpha \right\} \]
\[ d_2 = \text{tr}\left\{ (X'_\alpha X_\alpha)^{-1}X'_\alpha P_0 X_\alpha \right\} \]
\[ d_3 = \text{tr}\left\{ (X'_\alpha X_\alpha)^{-1}X'_\alpha P_0 X_\alpha \left( X'_\alpha X_\alpha \right)^{-1}X'_\alpha Z_0 Z'_0 X_\alpha \right\} \]

Wansbeek-Kapteyn Method

You can use the Wansbeek-Kapteyn method of estimating variance components by specifying the VCOMP=WK option in the MODEL statement. The method is a specialization (Baltagi and Chang 1994) of the approach used by Wansbeek and Kapteyn (1989) for unbalanced two-way models. The method was also suggested by Amemiya (1971) for balanced data.

The Wansbeek-Kapteyn method is an ANOVA method that uses the within residuals from one-way fixed effects, \( \hat{u}_w \), in both quadratic forms.

The expected values of the quadratic forms are

\[ E(\hat{u}_w'Q_0\hat{u}_w) = (M - N - K)\sigma_e^2 \]
\[ E(\hat{u}_w'P_0\hat{u}_w) = (N - 1 + d)\sigma_v^2 + \left( M - M^{-1} \sum_{i=1}^{N} T_i^2 \right) \sigma_e^2 \]

where

\[ d = \text{tr}\left\{ (X'Q_0X)^{-1}X'P_0X \right\} - \text{tr}\left\{ (X'Q_0X)^{-1}X'J_M X \right\} \]
Fuller-Battese Method

You can use the Fuller-Battese (1974) method of estimating variance components by specifying the VCOMP=FB option in the MODEL statement. Following Baltagi (2013, sec. 9.2), you obtain $\hat{\sigma}_v^2$ as the mean square error (MSE) from one-way fixed effects. The cross-sectional variance is

$$\hat{\sigma}_v^2 = \frac{R(v|\beta) - (N - 1)\hat{\sigma}_e^2}{M - \text{tr}\{Z_0X_\alpha(X_\alpha'X_\alpha)^{-1}X_\alpha'Z_0\}}$$

where

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

for

$$R(\nu) = y'Z_0(Z_0'Z_0)^{-1}Z_0'y$$
$$R(\beta|v) = y'_wX'_w(X'_wX_w)^{-1}X'_wy_w$$
$$R(\beta) = y'_\alpha X'_\alpha (X'_\alpha X_\alpha)^{-1}X'_\alpha y$$

Nerlove Method

You can use the Nerlove (1971) method of estimating variance components by specifying the VCOMP=NL option in the MODEL statement. The Nerlove method provides a simple alternative to the previous three estimation strategies. You estimate $\sigma_v^2$ as the sample variance of the cross-sectional effects, estimated from a one-way fixed-effects regression. Specifically, $\hat{\sigma}_v^2 = (N - 1)^{-1}\sum_{i=1}^N (\hat{v}_i - \hat{v})^2$, where $\hat{v}$ is the mean of the estimated fixed effects. You estimate $\sigma_e^2$ by taking the error sum of squares from one-way fixed-effects regression and then dividing by $M$.

Selecting the Appropriate Variance Component Method

By default, variance components are estimated by the Fuller-Battese method (VCOMP=FB) when the data are balanced, and by the Wansbeek-Kapteyn method (VCOMP=WK) when the data are unbalanced.

Baltagi and Chang (1994) conducted an extensive simulation study of the finite-sample properties of the variance estimators that the PANEL procedure supports. The choice of method has little bearing on estimates of regression coefficients, their standard errors, and estimation of the error variance $\sigma_e^2$. If your goal is inference on $\beta$, then the variance-component method will matter little.

The methods have varying performance in how they estimate $\sigma_v^2$, the cross-sectional variance. All four methods tend to perform poorly if either the data are severely unbalanced or the ratio $\sigma_v^2/\sigma_e^2$ is much greater than 1.

Of these four methods, the Nerlove method is the only one that guarantees a nonnegative estimate of $\sigma_v^2$; the other three methods reset a negative estimate to 0. However, the Nerlove method is particularly unsuitable for unbalanced data because the sample variance that it computes is not weighted by $T_i$. 
Two-Way Random-Effects Model (RANTWO Option)

You perform two-way random-effects estimation by specifying the RANTWO option in the MODEL statement (or by specifying nothing, because RANTWO is the default). The specification for the two-way random-effects model is

\[ u_{it} = v_i + \lambda_t + e_{it} \]

where the \( v_i \) are iid with zero mean and variance \( \sigma_v^2 \), the \( \lambda_t \) are iid with zero mean and variance \( \sigma_{\lambda}^2 \), and the \( e_{it} \) are iid with zero mean and variance \( \sigma_e^2 \). Furthermore, a random-effects specification assumes that the error terms are mutually uncorrelated and that each error term is uncorrelated with \( X \).

Estimation proceeds in two steps. First, you obtain estimates of the variance components \( \sigma_v^2, \sigma_{\lambda}^2, \) and \( \sigma_e^2 \). The PANEL procedure provides four methods of estimating variance components; these methods are described in the following subsections.

Second, with the variance-component estimates in hand, you transform the data in such a way that estimation can take place using ordinary least squares (OLS). In two-way models with unbalanced data, the transformation is quite complex. Throughout this section, \( y \) and \( X \) are treated as being sorted first by time, and then by cross section within time. For the definitions of the design matrices \( D_N \) and \( D_T \), see the section “Two-Way Fixed-Effects Model (FIXTWO Option)” on page 1841. The variance of \( y \) is

\[ \mathbf{\Omega} = \sigma_v^2 \mathbf{I}_N + \sigma_{\lambda}^2 \mathbf{D}_N \mathbf{D}_N' + \sigma_e^2 \mathbf{D}_T \mathbf{D}_T' \]

and estimation proceeds as OLS regression of \( \hat{\sigma}_e \hat{\Omega}^{-1/2} y \) on \( \hat{\sigma}_e \hat{\Omega}^{-1/2} X_{\alpha} \).

Rather than invert the \( M \times M \) matrix \( \hat{\Omega} \) directly, Wansbeek and Kapteyn (1989) provide the more convenient form

\[ \hat{\sigma}_e^2 \hat{\Omega}^{-1} = \mathbf{V} - \mathbf{V} \mathbf{D}_T \mathbf{P}^{-1} \mathbf{D}_T' \mathbf{V} \]

where

\[ \mathbf{V} = \mathbf{I}_M - \mathbf{D}_N \hat{\Delta}_N^{-1} \mathbf{D}_N' \]

\[ \hat{\Delta}_T = \mathbf{D}_T' \mathbf{D}_N \hat{\Delta}_N^{-1} \mathbf{D}_N' \mathbf{D}_T \]

with \( \hat{\Delta}_N = \mathbf{D}_N' \mathbf{D}_N + (\hat{\sigma}_v^2 / \hat{\sigma}_v^2) \mathbf{I}_N \) and \( \hat{\Delta}_T = \mathbf{D}_T' \mathbf{D}_T + (\hat{\sigma}_e^2 / \hat{\sigma}_e^2) \mathbf{I}_T \).

If the data are balanced, then the calculations are simplified considerably—the data are transformed from \( z_{it} \) to \( \hat{z}_i - \hat{\theta}_1 \hat{\lambda}_i, - \hat{\theta}_2 \hat{\lambda}_i, \) and \( \hat{\theta}_3 \hat{\lambda}_i \), where

\[ \hat{\theta}_1 = 1 - \hat{\sigma}_e (T \hat{\sigma}_v^2 + \hat{\sigma}_e^2)^{-1/2} \]

\[ \hat{\theta}_2 = 1 - \hat{\sigma}_e (N \hat{\sigma}_{\lambda}^2 + \hat{\sigma}_e^2)^{-1/2} \]

\[ \hat{\theta}_3 = \hat{\theta}_1 + \hat{\theta}_2 + \hat{\sigma}_e (T \hat{\sigma}_v^2 + N \hat{\sigma}_{\lambda}^2 + \hat{\sigma}_e^2)^{-1/2} - 1 \]

The PANEL procedure provides four methods of estimating variance components, as described in the following subsections.
Wallace-Hussain Method

You can use the Wallace-Hussain (1969) method of estimating variance components by specifying the VCOMP=WH option in the MODEL statement. The Wallace-Hussain method is part of a class of methods known as analysis of variance (ANOVA) estimators.

ANOVA estimators obtain variance components by solving a system of equations that is based on expected sums of squares. The following quadratic forms correspond to the two-way within sum of squares, the sum of squares between time periods, and the sum of squares between cross sections, respectively:

\[ q_e = u' Pu \]
\[ q_\lambda = u' D_T \Delta_T^{-1} D_T' u \]
\[ q_\nu = u' D_N \Delta_N^{-1} D_N' u \]

The matrix \( P \) is the two-way within transformation defined in the section “Two-Way Fixed-Effects Model (FXTWO Option)” on page 1841, \( \Delta_T = D_T D_T' \), \( \Delta_N = D_N D_N' \), and \( u \) is the vector of true residuals.

The ANOVA methods differ only in how they estimate \( u \). The Wallace-Hussain method is an ANOVA method that uses the residuals from pooled (OLS) regression, \( \hat{u}_p \), in all three quadratic forms.

The expected values of the quadratic forms are

\[ E \left( \hat{u}_p' P u \right) = d_{11} \sigma_e^2 + d_{12} \sigma_v^2 + d_{13} \sigma_\lambda^2 \]
\[ E \left( \hat{u}_p' P \lambda u \right) = d_{21} \sigma_e^2 + d_{22} \sigma_v^2 + d_{23} \sigma_\lambda^2 \]
\[ E \left( \hat{u}_p' P \nu u \right) = d_{31} \sigma_e^2 + d_{32} \sigma_v^2 + d_{33} \sigma_\lambda^2 \]

Define \( \Sigma = (X'_\alpha X_\alpha)^{-1} \), which is the inverse crossproducts matrix from pooled regression. Also define \( S_\nu = X'_\nu D_N D_N' X_\alpha \) and \( S_\lambda = X'_\lambda D_T D_T' X_\alpha \), which are the individual-level sum of squares and the time-level sum of squares, respectively. The coefficients are

\[ d_{11} = M - N - T + 1 \text{ tr } \left( X'_\alpha P X_\alpha \Sigma \right) \]
\[ d_{12} = \text{ tr } \left( S_\nu \Sigma X'_\alpha P X_\alpha \Sigma \right) \]
\[ d_{13} = \text{ tr } \left( S_\lambda \Sigma X'_\alpha P X_\alpha \Sigma \right) \]
\[ d_{21} = T - \text{ tr } \left( X'_\alpha P \lambda X_\alpha \Sigma \right) \]
\[ d_{22} = T - 2 \text{ tr } \left( X'_\alpha P \lambda D_N D_N' X_\alpha \Sigma \right) + \text{ tr } \left( X'_\alpha P \lambda X_\alpha \Sigma S_\nu \Sigma \right) \]
\[ d_{23} = M - 2 \text{ tr } (S_\lambda \Sigma) + \text{ tr } \left( X'_\alpha P \lambda X_\alpha \Sigma S_\lambda \Sigma \right) \]
\[ d_{31} = N - \text{ tr } \left( X'_\alpha P \nu X_\alpha \Sigma \right) \]
\[ d_{32} = M - 2 \text{ tr } (S_\nu \Sigma) + \text{ tr } \left( X'_\alpha P \nu X_\alpha \Sigma S_\nu \Sigma \right) \]
\[ d_{33} = N - 2 \text{ tr } \left( X'_\alpha P \nu D_T D_T' X_\alpha \Sigma \right) + \text{ tr } \left( X'_\alpha P \nu X_\alpha \Sigma S_\lambda \Sigma \right) \]
Wansbeek-Kapteyn Method

You can use the Wansbeek-Kapteyn method of estimating variance components by specifying the VCOMP=WK option in the MODEL statement. The method is a specialization (Baltagi and Chang 1994) of the approach used by Wansbeek and Kapteyn (1989) for unbalanced two-way models.

The Wansbeek-Kapteyn method is an ANOVA method that uses the within residuals from two-way fixed effects, \( \hat{u}_f \), in all three quadratic forms.

The expected values of the quadratic forms are

\[
E\left(\hat{u}^T_f P\hat{u}_f\right) = (M - N - T - K + 1)\sigma^2_e
\]

\[
E\left(\hat{u}^T_f P_\lambda \hat{u}_f\right) = (T + k_N - k_0)\sigma^2_e + (T - \delta_N)\sigma^2_\lambda + (M - \delta_T)\sigma^2_\lambda
\]

\[
E\left(\hat{u}^T_f P_v \hat{u}_f\right) = (N + k_T - k_0)\sigma^2_e + (M - \delta_N)\sigma^2_\nu + (N - \delta_T)\sigma^2_\lambda
\]

where \( \delta_N = M^{-1} \sum_{i=1}^N T_i^2 \) and \( \delta_T = M^{-1} \sum_{t=1}^T N_t^2 \). The other constants are defined by

\[
k_0 = 1 + M^{-1} j_M X^T (X'P_X)^{-1} X' j_M
\]

\[
k_N = \text{tr}\{X'P_X)^{-1} X' P_\lambda X\}
\]

\[
k_T = \text{tr}\{X'P_X)^{-1} X' P_v X\}
\]

When the NOINT option is specified, the variance-component equations change slightly: \( k_0, \delta_N, \) and \( \delta_T \) are all replaced by 0.

The Wansbeek-Kapteyn method is the default method when the data are unbalanced.

Fuller-Battese Method

You can use the Fuller-Battese (1974) method of estimating variance components by specifying the VCOMP=FB option in the MODEL statement. Following the discussion in Baltagi, Song, and Jung (2002), the Fuller-Battese method is a variation of the two ANOVA methods discussed previously in this section.

The quadratic form, \( q_\epsilon \), is the same as in the previous methods, and \( \hat{u} \) is estimated by the two-way within residuals \( \hat{u}_f \). The other two quadratic forms, \( q_\lambda \) and \( q_\nu \), are replaced by the error sums of squares from one-way fixed-effects estimations.

The resulting system of equations is

\[
E\left(\hat{u}^T_f P\hat{u}_f\right) = (M - N - T - K + 1)\sigma^2_e
\]

\[
E\left(\hat{u}^T_\lambda \hat{u}_\lambda\right) = (M - T - K)\sigma^2_e + \left[ M - T - \text{tr}\left\{X' W_\lambda D_N D'_N W_\lambda X \left(X' W_\lambda X\right)^{-1}\right\}\right] \sigma^2_\nu
\]

\[
E\left(\hat{u}^T_\nu \hat{u}_\nu\right) = (M - N - K)\sigma^2_e + \left[ M - N - \text{tr}\left\{X' W_\nu D_T D'_T W_\nu X \left(X' W_\nu X\right)^{-1}\right\}\right] \sigma^2_\lambda
\]

where \( W_\lambda = I_M - P_\lambda, W_\nu = I_M - P_\nu \). \( \hat{u}_\lambda \) are the residuals from a one-way model with time fixed effects, and \( \hat{u}_\nu \) are the residuals from a one-way model with individual fixed effects.

The Fuller-Battese method is the default method when the data are balanced.
Nerlove Method

You can use the Nerlove (1971) method of estimating variance components by specifying the VCOMP=NL option in the MODEL statement.

You begin by fitting a two-way fixed-effects model. The estimator of the error variance is

$$\hat{\sigma}_e^2 = M^{-1} \hat{\mathbf{u}}'_f \mathbf{P} \hat{\mathbf{u}}_f$$

You obtain $\hat{\sigma}_e^2$ as the sample variance of the $N$ estimated individual effects, and $\hat{\sigma}_\lambda^2$ as the sample variance of the $T$ estimated time effects.

Parks Method for Autoregressive Models (PARKS Option)

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

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

where

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

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 $\mathbf{u}$ can be expressed as

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

where

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

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

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

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

which is written

\[
y_{it} = \sum_{k=1}^{p} X_{i1k} \hat{\beta}_k + u_{it} - \hat{\rho}_i u_{it-1} t = 2, \ldots, T
\]

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

The second step in estimating the covariance matrix \( V \) is 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

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

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,

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

where \( \hat{\Sigma} \) is the derived consistent estimator of \( V \). 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 $\hat{\Phi} = [\hat{\phi}_{ij}]_{i,j=1,...,N}$.

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

The variance estimate is

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

**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:

$$r_i = \begin{cases} 
  r_i & \text{if } |r_i| < 1 \\
  \max(0.95, r_{\text{max}}) & \text{if } r_i \geq 1 \\
  \min(-0.95, r_{\text{min}}) & \text{if } r_i \leq -1 
\end{cases}$$

where

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

and

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

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

---

**Da Silva Method for Moving Average Models (DASILVA Option)**

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 $1_N$ 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 $\mathbb{R}^p$ whose elements are uniformly bounded in $\mathbb{R}^p$. 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. $e_{it} = (e_{i1}, \ldots, e_{iT})'$ is a sample of a realization of a finite moving-average time series of order $m < T - 1$ for each $i$; hence,

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

where $\alpha_0, \alpha_1, \ldots, \alpha_m$ are unknown constants such that $\alpha_0 \neq 0$ and $\alpha_m \neq 0$, and $\{e_{ij}\}_{j=-\infty}^{i=\infty}$ 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 $\sigma_e^2 > 0$. $\{e_{ij}\}_{j=-\infty}^{i=\infty}$ 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 $e_{t-k} \sim N(0, \sigma_e^2)$, 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

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

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

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

where \( \psi(k) = \sigma^2_x \sum_{j=0}^{m-k} \alpha_j \alpha_{j+k} \) for \( k = |t-s| \). For the definition of \( I_N, I_T, J_N, \) and \( J_T \), see the section “Fuller-Battese Method” on page 1845.

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

\[ V = \sigma^2_a (I_N \otimes J_T) + \sigma^2_b (J_N \otimes I_T) + \sum_{k=0}^{m} \psi(k)(I_N \otimes \Psi_T^{(k)}) \]

where \( \Psi_T^{(0)} = I_T \), and, for \( k = 1, \ldots, m \), \( \Psi_T^{(k)} \) is a band matrix whose \( k \)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

\[ v_1 = \sigma^2_a \]
\[ v_2 = \sigma^2_b \]
\[ v_k = \psi(k-3)k = 3, \ldots, m + 3 \]
\[ V_1 = I_N \otimes J_T \]
\[ V_2 = J_N \otimes I_T \]
\[ V_k = I_N \otimes \Psi_T^{(k-3)}k = 3, \ldots, m + 3 \]

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 \( V_i \) 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).
**Hausman-Taylor Estimation (HTAYLOR Option)**

You perform Hausman-Taylor estimation by specifying the HTAYLOR option in the MODEL statement. The Hausman and Taylor (1981) model is a hybrid that combines the consistency of a fixed-effects model with the efficiency and applicability of a random-effects model. One-way random-effects models assume exogeneity of the regressors; that is, they are independent of both the cross-sectional and observation-level errors. When some regressors are correlated with the cross-sectional errors, you can adjust the random-effects model to deal with this form of endogeneity.

Consider the one-way model:

\[ y_{it} = x_{1it}\beta_1 + x_{2it}\beta_2 + z_{1i}\gamma_1 + z_{2i}\gamma_2 + \nu_i + \epsilon_{it} \]

The regressors are subdivided so that \( x_{1it} \) and \( x_{2it} \) vary within cross sections, whereas \( z_{1i} \) and \( z_{2i} \) do not and would otherwise be dropped from a fixed-effects model. The subscript 1 denotes variables that are independent of both error terms (exogenous variables), and the subscript 2 denotes variables that are independent of the observation-level errors \( \epsilon_i \) but correlated with cross-sectional errors \( \nu_i \). The intercept term (if your model has one) is included as part of \( z_{1i} \).

The Hausman-Taylor estimator is a two-stage least squares (2SLS) regression on data that are weighted similarly to data for random-effects estimation. The weights are functions of the estimated variance components.

The observation-level variance is estimated from a one-way fixed-effects model fit. Obtain \( y_w \), \( X_w \), and \( \hat{\beta}_w \) from the section “One-Way Fixed-Effects Model (FIXONE and FIXONETIME Options)” on page 1840. Then \( \hat{\sigma}_e^2 = \text{SSE}/(M - N) \), where

\[ \text{SSE} = \left( y_w - X_w\hat{\beta}_w \right)\left( y_w - X_w\hat{\beta}_w \right)' \]

To estimate the cross-sectional error variance, form the mean-residual vector \( r = P_0'(y - X_w\hat{\beta}_w) \), where \( P_0 = \text{diag}(J_{T_i}) \). You can use the mean residuals to obtain intermediate estimates of the coefficients for \( z_1 \) and \( z_2 \) via two-stage least squares (2SLS) estimation. At the first stage, use \( x_1 \) and \( z_1 \) as instrumental variables to predict \( z_2 \). At the second stage, regress \( r \) on both \( z_1 \) and the predicted \( z_2 \) to obtain \( \hat{\gamma}_1^m \) and \( \hat{\gamma}_2^m \).

To estimate the cross-sectional variance, compute \( \hat{\sigma}_v^2 = \{R(v)/N - \hat{\sigma}_e^2\}/\hat{\bar{T}} \), where \( \bar{T} = N/(\sum_{i=1}^N T_i^{-1}) \) and

\[ R(v) = (r - Z_1\hat{\gamma}_1^m - Z_2\hat{\gamma}_2^m)'(r - Z_1\hat{\gamma}_1^m - Z_2\hat{\gamma}_2^m) \]

The design matrices \( Z_1 \) and \( Z_2 \) are formed by stacking the data observations of \( z_{1i} \) and \( z_{2i} \), respectively.

After variance-component estimation, transform the dependent variable into partial deviations: \( y_{it}^* = y_{it} - \hat{\bar{T}}\hat{\gamma}_i \). Likewise, transform the regressors to form \( x_{1it}^* \), \( x_{2it}^* \), \( z_{1i}^* \), and \( z_{2i}^* \). The partial weights \( \theta_i \) are determined by \( \theta_i = 1 - \hat{\sigma}_v / w_i \), with \( w_i^2 = T_i\hat{\sigma}_v^2 + \hat{\sigma}_e^2 \).

Finally, you obtain the Hausman-Taylor estimates by performing 2SLS regression of \( y_{it}^* \) on \( x_{1it}^* \), \( x_{2it}^* \), \( z_{1i}^* \), and \( z_{2i}^* \). For the first-stage regression, use the following instruments:

- \( \bar{x}_{it} \), the deviations from cross-sectional means for all time-varying variables (correlated and uncorrelated) for the \( i \)th cross section during time period \( t \)
- \( (1 - \theta_i)\bar{x}_{1i} \), where \( \bar{x}_{1i} \) are the means of the time-varying exogenous variables for the \( i \)th cross section
Multiplication by the factor \((1 - \hat{\theta}_i)\) is redundant in balanced data but necessary in the unbalanced case to produce accurate instrumentation; see Gardner (1998).

Let \(k_1\) equal the number of regressors in \(x_1\), and let \(g_2\) equal the number of regressors in \(z_2\). Then the Hausman-Taylor model is identified only if \(k_1 \geq g_2\); otherwise, no estimation takes place.

Hausman and Taylor (1981) describe a specification test that compares their model to a fixed-effects model. For a null hypothesis of fixed effects, Hausman’s \(m\) statistic is calculated by comparing the parameter estimates and variance matrices for both models, which is identical to how it is calculated for one-way random-effects models; for more information, see the section “Hausman Test” on page 1873. However, the number of degrees of freedom of the test is not based on matrix rank but instead is equal to \(k_1 - g_2\).

**Amemiya-MaCurdy Estimation (AMACURDY Option)**

You perform Amemiya-MaCurdy estimation by specifying the AMACURDY option in the MODEL statement. The Amemiya-MaCurdy (1986) model is similar to the Hausman-Taylor model. Following the development in the section “Hausman-Taylor Estimation (HTAYLOR Option)” on page 1854, estimation is identical up to the final 2SLS instrumental variables regression. In addition to the set of instruments that the Hausman-Taylor estimator uses, you use the following:

\[
\mathbf{x}_{1i1}, \mathbf{x}_{1i2}, \ldots, \mathbf{x}_{1iT}
\]

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

The Amemiya-MaCurdy model attempts to gain efficiency over the Hausman-Taylor model by adding instruments. This comes at a price of a more stringent assumption on the exogeneity of the \(x_1\) variables. Although the Hausman-Taylor model requires only that the cross-sectional means of \(x_1\) be orthogonal to \(\mathbf{v}_i\), the Amemiya-MaCurdy estimation requires orthogonality at every point in time; see Baltagi (2013, sec. 7.4).

A Hausman specification test is provided to test the validity of the added assumption. Define \(\alpha' = (\beta_1', \beta_2', y_1', y_2')\), its Hausman-Taylor estimate as \(\hat{\alpha}_{HT}\), and its Amemiya-MaCurdy estimate as \(\hat{\alpha}_{AM}\). Under the null hypothesis, both estimators are consistent and \(\hat{\alpha}_{AM}\) is efficient. The Hausman test statistic is

\[
m = (\hat{\alpha}_{HT} - \hat{\alpha}_{AM})' \left( \hat{\Sigma}_{HT} - \hat{\Sigma}_{AM} \right)^{-1} (\hat{\alpha}_{HT} - \hat{\alpha}_{AM})
\]

where \(\hat{\Sigma}_{HT}\) and \(\hat{\Sigma}_{AM}\) are variance-covariance estimates of \(\hat{\alpha}_{HT}\) and \(\hat{\alpha}_{AM}\), respectively. Under the null hypothesis, \(m\) follows a \(\chi^2\) distributed with degrees of freedom equal to the rank of \((\hat{\Sigma}_{HT} - \hat{\Sigma}_{AM})^{-1}\).

**Dynamic Panel Estimation (DYNDIFF and DYNSYS Options)**

You perform dynamic panel estimation that uses first differences by specifying the DYNDIFF option in the MODEL statement. For dynamic panel estimation that uses a full system of difference and level equations, specify the DYNSYS option. For an example of dynamic panel estimation, see Example 26.4.
Dynamic panel models are regression models that include lagged versions of the dependent variable as covariates. Consider the following panel regression, which includes $L$ lags of the dependent variable:

$$y_{it} = \sum_{j=1}^{L} \phi_j y_{i, t-j} + \sum_{k=1}^{K} x_{itk} \beta_k + v_i + \epsilon_{it}$$

Because the effect $v_i$ is common to all observations for that individual, it is correlated with any lagged $y$ because it played a role in its realization. As such, lagged dependent variables are endogenous regressors and require special consideration.

**First Differencing**

For ease of notation, consider the special case $L = K = 1$. A first attempt to remove the source of the correlation would be to take first differences, which removes $v_i$. That is,

$$\Delta y_{it} = \Delta y_{i, t-1} + \Delta x_{it} \beta + \eta_{it}$$

where $\Delta y_{it} = y_{i,t} - y_{i,t-1}$, $\Delta x_{it} = x_{i,t} - x_{i,t-1}$, and $\eta_{it} = \epsilon_{i,t} - \epsilon_{i,t-1}$. Even though the individual effects are removed, the problem of endogeneity persists because $\Delta y_{i,t-1}$ is correlated with the differenced error term $\eta_{it}$. That is because $\epsilon_{i,t-1}$ is a component of $y_{i,t-1}$ (Nickell 1981).

Arellano and Bond (1991) show that you can use the generalized method of moments (GMM) to obtain a consistent estimator. In GMM parlance, the moment condition that $\mathbb{E}(\Delta y_{i,t-1} \eta_{it}) = 0$ is violated. Estimation requires a set of instrumental variables that do meet their moment conditions and that can adequately predict $\Delta y_{i,t-1}$. A natural set of instruments is $y_{i,t-2}$ and all other previous realizations of $y$. These lags of $y$ are not correlated with $\epsilon_{i,t-1}$ because they occurred before time $t - 1$. Given the autoregressive nature of the model, $y_{i,t-1}$ (and hence $\Delta y_{i,t-1}$) is well predicted by its previous values.

Begin with $t = 3$, the first time period where the differenced model holds. The dynamic regression model for individual $i$ can be expressed as

$$y_{i,t}^d = X_{i,t}^d \gamma + \eta_{i,t}^d$$

where

$$y_{i,t}^d = \begin{pmatrix} \Delta y_{i3} \\ \Delta y_{i4} \\ \vdots \\ \Delta y_{iT} \end{pmatrix}, \quad X_{i,t}^d = \begin{pmatrix} \Delta y_{i2} & \Delta x_{i3} \\ \Delta y_{i3} & \Delta x_{i4} \\ \vdots & \vdots \\ \Delta y_{iT-1} & \Delta x_{iT} \end{pmatrix}, \quad \gamma = \begin{pmatrix} \phi \\ \beta \end{pmatrix}, \quad \eta_{i,t}^d = \begin{pmatrix} \eta_{i3} \\ \eta_{i4} \\ \vdots \\ \eta_{iT} \end{pmatrix}$$

Proceeding with the idea that you can use $(y_{i1}, \ldots, y_{i, t-1})$ as instruments for $\Delta y_{i,t}$, the instrument matrix for the lagged dependent variables is

$$Z_{i,t}^d = \begin{pmatrix} y_{i1} & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & y_{i1} & y_{i2} & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & y_{i1} & y_{i2} & y_{i3} & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & y_{i1} & \cdots & y_{i, T-2} \end{pmatrix}$$

This extends naturally to $L > 1$ and $K > 1$; simply add columns to $X_{i,t}^d$ and elements to $\gamma$ as appropriate.

When an observation is either missing or lost because of missing lags, delete the corresponding rows of $y_{i,t}^d$, and so on.
Dynamic Panel Estimation (DYNDIFF and DYN SYS Options)

\[ X_i^d, \eta_i^d, \text{ and } Z_i^d. \] Even if an observation is not missing with respect to the regression model, some of the lagged instruments might not be available because previous observations are missing. When that occurs, replace any missing instrument with 0.

When you specify the DYNDIFF option in the MODEL statement, PROC PANEL by default treats \( x \) variables as exogenous and uses a projection that leaves these variables unchanged in the differenced regression. The full instrument matrix is then \( Z_i = (Z_i^d, D_i) \), where

\[
D_i = \begin{pmatrix}
\Delta x_{i31} & \Delta x_{i32} & \cdots & \Delta x_{i3K} \\
\Delta x_{i41} & \Delta x_{i42} & \cdots & \Delta x_{i4K} \\
\vdots & \vdots & \ddots & \vdots \\
\Delta x_{iT1} & \Delta x_{iT2} & \cdots & \Delta x_{iTK}
\end{pmatrix}
\]

When \( L = 1 \), the default \( Z_i \) has \( (T-1)(T-2)/2 + K \) columns. Each column \( z_c \) of \( Z_i \) satisfies the moment condition \( E(z_c^\prime \eta_i^d) = 0 \).

**System GMM**

Blundell and Bond (1998) proposed a system GMM estimator that uses additional moment conditions to increase efficiency. The efficiency gain can be substantial when there is strong serial correlation in the dependent variable.

When either \( \phi \) is near 1 or \( \alpha_i^2 / \sigma_i^2 \) is large, the lagged dependent variables \( y_{it-1} \) are weak instruments for the differenced variables \( \Delta y_{it} \). System GMM solves the weak instrument problem by augmenting the difference equations described previously with a set of level equations. When \( L = K = 1 \), the level equations are

\[
y_i^l = X_i^l \gamma + \epsilon_i^l
\]

where

\[
y_i^l = \begin{pmatrix} y_{i2} \\ y_{i3} \\ \vdots \\ y_{iT-1} \\ y_{iT} \end{pmatrix}, \quad X_i^l = \begin{pmatrix} y_{i1} & x_{i2} \\ y_{i2} & x_{i3} \\ \vdots & \vdots \\ y_{i,T-1} & x_{iT} \end{pmatrix}, \quad \epsilon_i^l = \begin{pmatrix} \nu_i + \epsilon_{i2} \\ \nu_i + \epsilon_{i3} \\ \vdots \\ \nu_i + \epsilon_{iT} \end{pmatrix}
\]

Blundell and Bond (1998) note that you can use lagged differences of \( y \) as instruments for the levels of \( y \). The main instrument matrix for the level equations is then

\[
Z_i^l = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & \Delta y_{i2} & 0 & \cdots & 0 \\ 0 & 0 & \Delta y_{i3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \Delta y_{iT-1} \\ \end{pmatrix}
\]

where the first row corresponds to time \( t = 2 \). You can extend this to \( L > 1 \) and \( K > 1 \) by adding columns to \( X_i^l \) and elements to \( \gamma \) as appropriate. Higher-order lags require deletion of the leading rows of \( y_i^l, X_i^l, \epsilon_i^l \), and \( Z_i^l \).

Regression on the full system is obtained by stacking \( y_i^d \) and \( y_i^l \) to form \( y_i \), stacking \( X_i^d \) and \( X_i^l \) to form \( X_i \), and stacking \( \eta_i^d \) and \( \epsilon_i^l \) to form \( \epsilon_i \).
When you specify the DYNSYS model option, the default instrument matrix for the full system is

$$Z_i = \left( \begin{array}{ccc} Z_i^d & 0 & D_i \\ 0 & Z_i^e & 0 \end{array} \right)$$

**Estimation**

The estimation in this section assumes system GMM. To obtain difference GMM, restrict estimation to the rows that correspond to the difference equations.

The initial moment matrix is derived from the theoretical variance of the combined residuals and is expressed as $H_{1i} = \text{diag}(G_{1i}, G_{2i})$, where

$$G_{1i} = \begin{pmatrix} 1 & -0.5 & 0 & \cdots & 0 & 0 & 0 \\ -0.5 & 1 & -0.5 & \cdots & 0 & 0 & 0 \\ 0 & -0.5 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -0.5 & 0 \\ 0 & 0 & 0 & \cdots & -0.5 & 1 & -0.5 \\ 0 & 0 & 0 & \cdots & 0 & -0.5 & 1 \end{pmatrix}$$

and $G_{2i}$ is 0.5 times the identity matrix.

Define the weighting matrix as

$$W_1 = \left( \sum_{i=1}^{N} Z_i'H_{1i}Z_i \right)^{-1}$$

and the projections as

$$P_y = \sum_{i=1}^{N} Z_i'y_i; \quad P_x = \sum_{i=1}^{N} Z_i'X_i$$

The one-step GMM estimate of $y$ is the weighted OLS estimator

$$\hat{y}_1 = \left( P_x'W_1P_x \right)^{-1} P_x'W_1P_y$$

The variance of $\hat{y}_1$ is

$$\text{Var} (\hat{y}_1) = \hat{\sigma}_e^2 \left( P_x'W_1P_x \right)^{-1}$$

where $\hat{\sigma}_e^2$ is the mean square error (MSE) derived solely from the difference equations, namely

$$\hat{\sigma}_e^2 = (M - K)^{-1} \sum_{i=1}^{N} \left( y_i^d - X_i^d \hat{y}_1 \right) \left( y_i^d - X_i^d \hat{y}_1 \right)'$$

The total number of observations, $M$, is equal to the number of observations for which the difference equations hold.
A disadvantage of $\hat{\gamma}_1$ is its reliance on the theoretical basis of $H_{1i}$. The two-step GMM estimate of $\gamma$ replaces $H_{1i}$ with a version that is obtained from the observed one-step residuals. Let $H_{2i}$ be the outer product of $\hat{\epsilon}_i^s = y_i^s - X_i^s \hat{\gamma}_1$. Then

$$\hat{\gamma}_2 = \left( P_x' W_2 P_x \right)^{-1} P_x' W_2 P_y$$

where

$$W_2 = \left( \sum_{i=1}^{N} Z_i' H_{2i} Z_i \right)^{-1}$$

The variance of $\hat{\gamma}_2$ is

$$\text{Var} (\hat{\gamma}_2) = \left( P_x' W_2 P_x \right)^{-1}$$

The iterated GMM estimator of $\gamma$ continues this pattern: First, use the current estimate $\hat{\gamma}_c$ to form the residuals that compose $H_{c+1,i}$. Second, use $H_{c+1,i}$ to form the weighting matrix $W_{c+1}$. Third, use $W_{c+1}$ to update the estimate $\hat{\gamma}_{c+1}$.

There are two criteria by which convergence is achieved. The first (and default) criterion is met when the magnitude of $\hat{\gamma}_c$ changes by a relative amount smaller than $b$, as specified in the BTOL= option in the MODEL statement. The second criterion is met when the magnitude of the variance matrix changes by a relative amount smaller than $a$, as specified in the ATOL= option in the MODEL statement.

Robust variances are calculated by the sandwich method. The robust variance of $\hat{\gamma}_1$ is

$$\text{Var}^r (\hat{\gamma}_1) = \left( P_x' W_1 P_x \right)^{-1} P_x' W_1 W_2^{-1} W_1 P_x \left( P_x' W_1 P_x \right)^{-1}$$

The robust variance of $\hat{\gamma}_2$ is

$$\text{Var}^r (\hat{\gamma}_2) = \left( P_x' W_2 P_x \right)^{-1} P_x' W_2 W_3^{-1} W_2 P_x \left( P_x' W_2 P_x \right)^{-1}$$

and so on as you iterate $\hat{\gamma}_c$.

Arellano and Bond (1991), among others, note that robust two-step variance estimators are biased. Windmeijer (2005) derived a bias-corrected variance of $\hat{\gamma}_2$, and you can obtain this correction by specifying the BIASCORRECTED option in the MODEL statement.

Define the one-step and two-step residuals as $\hat{\epsilon}_{1i} = y_i^s - X_i^s \hat{\gamma}_1$ and $\hat{\epsilon}_{2i} = y_i^s - X_i^s \hat{\gamma}_2$. Also define the projected two-step residual as

$$P_e = \sum_{i=1}^{N} Z_i' \hat{\epsilon}_{2i}$$

Formulate the matrix $D$ such that its $k$th column is $D_k = V_2 P_x' W_2 F_k W_2 P_e$, where $V_2 = \text{Var}(\hat{\gamma}_2)$. The matrix $F_k$ is the quadratic form

$$F_k = \sum_{i=1}^{N} Z_i' \left( x_{ik} \hat{\epsilon}_{1i} + \hat{\epsilon}_{1i} x_{ik}' \right) Z_i$$
where \( x_{ik} \) is the \( k \)th column of \( X^i \).

The Windmeijer (2005) bias-corrected variance is

\[
\text{Var}^w(\hat{y}_2) = V_2 + D V_2 D' + D V_1 D'
\]

where \( V^*_1 \) is the robust variance estimate of \( \hat{y}_1 \).

### Estimating the Intercept

The intercept term vanishes when you take first differences and is thus identified only in the level equations. If you specify the DYNDIFF option in the MODEL statement and your model includes an intercept, then PROC PANEL will fit the model by using system GMM with the following (default) instrumentation,

\[
Z_i = \begin{pmatrix} Z^d_i & D_i & 0 \\ 0 & 0 & j_i \end{pmatrix}
\]

where \( j_i \) is a column of ones. Because all the level instruments are zero except the constant, parameter estimates other than the intercept are unaffected by the added level equations.

If you specify the DYNDIFF option in the MODEL statement and your model does not include an intercept, then the level equations are excluded from the estimation.

If you specify the DYNSYS option in the MODEL statement, then there is no issue regarding the intercept. Under the default instrument specification, if \( X^i \) includes an intercept, then the level instruments include an added column of ones. That is,

\[
Z_i = \begin{pmatrix} Z^d_i & 0 & D_i & 0 \\ 0 & Z^i & 0 & j_i \end{pmatrix}
\]

### Customizing Instruments

When you specify the DYNSYS option for performing system GMM, the default instrument matrix is

\[
Z_i = \begin{pmatrix} Z^d_i & 0 & D_i & 0 \\ 0 & Z^i & 0 & c_i \end{pmatrix}
\]

where \( c_i \) is either a column of ones, or 0 if you specify the NOINT option.

You can override the default set of instruments by specifying an INSTRUMENTS statement. You can choose which instrument sets to include as components of \( Z_i \). The INSTRUMENTS statement provides options to generate the appropriate instruments when variables are either endogenous, predetermined, or exogenous.

The following discussion assumes that you are performing system GMM by using the DYNSYS option in the MODEL statement. When you specify the DYNDIFF option instead, any specification (except the constant \( c_i \)) that pertains to the level equations is ignored.

### Dependent Variable

The DEPVAR option in the INSTRUMENTS statement adds instruments for the dependent variable and its lags. Specifying DEPVAR(DIFF) includes the lagged levels of the dependent variable (the matrix \( Z^d_i \)) in the difference equations. Specifying DEPVAR(LEVEL) includes the first differences of the dependent variable.
(the matrix $Z_i^d$) in the level equations. Specifying DEPVAR(BOTH) (or simply DEPVAR) includes both $Z_i^d$ and $Z_i^\ell$.

You should at a minimum include instruments for the dependent variable when you perform dynamic panel estimation. For example:

```plaintext
proc panel data=a;
  id State Year;
  instruments depvar;
  model Sales = Price PopDensity / dynsys;
run;
```

**Constant (or Intercept)**

Specifying the keyword CONSTANT includes the constant vector $e_i$ in the level equations.

**Endogenous Variables**

A variable $x_{it}$ is endogenous if $E(x_{it}e_{is}) \neq 0$ for $s \leq t$ and 0 otherwise.

The DIFFEND= option specifies a list of endogenous variables that form instrument matrices for the difference equations. The instruments are “GMM-style” and mirror the form used for the dependent variable. Suppose that the model includes one lag of the dependent variable ($L = 1$). Specifying DIFFEND=(X) adds the following instruments to the difference equations:

$$G_i^d = \begin{pmatrix} x_{i1} & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & x_{i1} & x_{i2} & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & 0 & x_{i1} & x_{i2} & x_{i3} & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & x_{i1} & \cdots & x_i, T-2 \end{pmatrix}$$

The first row corresponds to time $t = 3$. The instruments are in lagged levels.

The LEVELEND= option specifies a list of endogenous variables that form instrument matrices for the level equations. The instruments mirror the form used for the dependent variable. Suppose that the model includes one lag of the dependent variable ($L = 1$). Specifying LEVELEND=(X) adds the following instruments to the level equations:

$$G_i^\ell = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 \\
0 & \Delta x_{i2} & 0 & \cdots & 0 \\
0 & 0 & \Delta x_{i3} & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \Delta x_i, T-1 \end{pmatrix}$$

The first row corresponds to time $t = 2$. Because the instruments are used for the level equations, they are in lagged differences.

The following code fits a dynamic panel model by using difference equations. It includes GMM-style instruments for both the dependent variable Sales and the variable Price:
proc panel data=a;
   id State Year;
   instruments depvar diffend = (Price);
   model Sales = Price PopDensity / dyndiff;
run;

Predetermined Variables

A variable $x_{it}$ is predetermined if $E(x_{it}|\epsilon_{is}) \neq 0$ for $s < t$ and 0 otherwise.

The DIFFPRE= option specifies a list of variables that are considered to be predetermined in the difference equations. The DIFFPRE= option works similarly to the DIFFEND= option, except that each observation contains an extra instrument that reflects orthogonality in the current time period. If $L = 1$, specifying DIFFPRE=(X) adds the following instruments to the difference equations:

$$P_i^d = \begin{pmatrix} x_{i1} & x_{i2} & 0 & 0 & 0 & \cdots & 0 & 0 & 0 \\
0 & 0 & x_{i1} & x_{i2} & x_{i3} & \cdots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & 0 & x_{i1} & \cdots & x_{i,T-1} \end{pmatrix}$$

The first row corresponds to time $t = 3$.

The LEVELPRE= option specifies a list of variables that are considered to be predetermined in the level equations. The LEVELPRE= option works similarly to the LEVELEND= option, except that the lag is shifted up to reflect orthogonality in the current time period. If $L = 1$, specifying LEVELPRE=(X) adds the following instruments to the level equations:

$$P_i^l = \begin{pmatrix} \Delta x_{i2} & 0 & 0 & \cdots & 0 \\
0 & \Delta x_{i3} & 0 & \cdots & 0 \\
0 & 0 & \Delta x_{i4} & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \Delta x_{i,T} \end{pmatrix}$$

The first row corresponds to time $t = 2$.

The following code fits a dynamic panel model by using difference equations. The instrument set includes GMM-style instruments for the dependent variable Sales and GMM-style instruments that correspond to the predetermined variable Price:

proc panel data=a;
   id State Year;
   instruments depvar diffpre = (Price);
   model Sales = Price PopDensity / dyndiff;
run;

Exogenous Variables

Exogenous variables are uncorrelated with both the level residuals and the differenced residuals. If a regression variable is exogenous, you might want to include that variable in the instrument set as a standard instrument. The DIFFEQ= option specifies a list of variables that compose the matrix of standard instruments.
**Dynamic Panel Estimation (DYNDIFF and DYNYSYS Options)**

\( D_i \) for the difference equations; for an example of how \( D_i \) is formed, see the section “First Differencing” on page 1856. These variables are usually exogenous regressors that you want to preserve under the projection to the instrument space. Because these instruments belong to the difference equations, the variables are automatically differenced.

The LEVELEQ= option specifies a list of variables that form a matrix of standard instruments that is included in the level equations. You can use this option to specify external instruments that are not part of the main regression but that can be used as instruments for the regression variables in levels.

If \( L = 1 \), specifying LEVELEQ=(X1 X2) adds the following instruments to the level equations:

\[
L_t = \begin{pmatrix}
  x_{i11} & x_{i21} \\
  x_{i12} & x_{i22} \\
  \vdots & \vdots \\
  x_{iT1} & x_{iT2}
\end{pmatrix}
\]

The first row corresponds to time \( t = 2 \).

The following example illustrates how you would use an INSTRUMENTS statement to obtain the default set of instruments for system GMM:

```plaintext
proc panel data=a;
   id State Year;
   instruments depvar(both) constant diffeq = (Price PopDensity);
   model Sales = Price PopDensity / dynsys;
run;
```

**Limiting the Number of Instruments**

Arellano and Bond’s (1991) technique of expanding instruments is a useful method of dealing with autocorrelation in the response variable. However, too many instruments can bias the estimator. The number of instruments grows quadratically with the number of time periods, making computations less feasible for larger \( T \).

By default, PROC PANEL uses all available lags. You can limit the number of instruments by specifying the MAXBAND= option in the INSTRUMENTS statement. For example, specifying MAXBAND=5 limits the number of GMM-style instruments to five per observation, for each variable. The MAXBAND= option applies to all GMM-style instruments: those for the dependent variable, those from the DIFFEND= option, and those from the DIFFPRE= option.

**Sargan Test of Overidentifying Restrictions**

A Sargan test is a referendum on your choice of instruments in a dynamic panel model. The Sargan test statistic for one-step GMM is

\[
J = \frac{1}{\hat{\sigma}^2} \left( \sum_{i=1}^{N} Z_i' \hat{e}_{1i} \right) W_1 \left( \sum_{i=1}^{N} Z_i' \hat{e}_{1i} \right)
\]

The Sargan test statistic for two-step GMM is

\[
J = \left( \sum_{i=1}^{N} Z_i' \hat{e}_{2i} \right) W_2 \left( \sum_{i=1}^{N} Z_i' \hat{e}_{2i} \right)
\]
It is similarly incremented for further iterations of GMM.

The null hypothesis of the Sargan test is that the moment conditions (as defined by the columns $Z_i$) hold, and thus $Z_i$ form an adequate set of instruments. Under the null, $J$ is distributed as $\chi^2$ with degrees of freedom equal to the rank of $W_c$ minus the number of parameters $K$. The nominal rank of $W_c$ is equal to the number of instruments. However, this number can be reduced because of collinearity and redundancy in the instrument specification. Furthermore, when $c > 1$, the maximum rank of $W_c$ is $N$, regardless of the number of instruments.

You should treat Sargan tests with caution when robust variances are used in the estimation. The theoretical distribution of $J$ does not hold under conditions that favor robust variances.

**AR($m$) Tests**

An AR($m$) test is a test for autocorrelation of order $m$ in the model residuals. Let $R_i^x$ be the working variance of the residuals from the full system. The precise definition of $R_i^x$ depends on the GMM stage and whether robust variances are specified; see Table 26.3.

**Table 26.3** Definition of the Working Residual Variance

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$R_i^x$</th>
</tr>
</thead>
<tbody>
<tr>
<td>One-step</td>
<td>$\hat{\sigma}<em>c^2 H</em>{1i}$</td>
</tr>
<tr>
<td>One-step, robust</td>
<td>$H_{2i}$</td>
</tr>
<tr>
<td>Two-step</td>
<td>$H_{2i}$</td>
</tr>
<tr>
<td>Two-step, robust</td>
<td>$H_{3i}$</td>
</tr>
<tr>
<td>Iteration $c$</td>
<td>$H_{ci}$</td>
</tr>
<tr>
<td>Iteration $c$, robust</td>
<td>$H_{c+1,i}$</td>
</tr>
</tbody>
</table>

Define the residual vector

$$\hat{e}_i = \begin{pmatrix} \hat{\eta}_i^d \\ 0 \end{pmatrix}$$

where $\hat{\eta}_i^d = y_i^d - X_i^d \hat{\gamma}_c$ are the residuals from the difference equations, evaluated at the final estimate of $\hat{\gamma}_c$. The trailing zeros correspond to the level equations. Define $\hat{\omega}_{mi}$ as a lagged version of $\hat{e}_i$ such that the following are true:

1. The first $m$ elements of $\hat{\omega}_{mi}$ are 0.
2. The next $p - m$ elements of $\hat{\omega}_{mi}$ are the first $p - m$ elements of $\hat{e}_i$, where $p$ is the number of difference equations.
3. The trailing elements of $\hat{\omega}_{mi}$ that correspond to the level equations are 0.

Define the following:

$$P_m = \sum_{i=1}^{N} Z_i' R_i^x \hat{\omega}_{mi}$$

$$Q_m = \sum_{i=1}^{N} \hat{\omega}_{mi}' X_i^x$$
The AR($m$) test statistic is $Z_m = k_{0m} (k_{1m} + k_{2m} + k_{3m})^{-1/2}$, where

\[
\begin{align*}
    k_{0m} &= \sum_{i=1}^{N} \hat{\omega}_{mi} \hat{e}_i \\
    k_{1m} &= \sum_{i=1}^{N} \hat{\omega}_{mi} R\hat{i}_{mi} \\
    k_{2m} &= -2Q_m (P'_x W_c P_x)^{-1} P'_x W_c P_m \\
    k_{3m} &= Q_m VQ_m
\end{align*}
\]

The matrix $V$ is the estimated variance matrix of the parameters, corresponding to the GMM stage specified, and either model-based, robust, or bias-corrected.

Under the null hypothesis of no autocorrelation, $Z_m$ follows a standard normal distribution. Because of the differencing in the errors, well-specified models present autocorrelation of order $m = 1$, but any autocorrelation at higher orders indicates a violation of assumptions.

---

**Restricted Estimation**

The PANEL procedure can fit models that have linear restrictions, producing a Lagrange multiplier (LM) test for each restriction. Consider a set of $J$ linear restrictions $R\beta = q$, where $R$ is $J \times K$ and $q$ is $J \times 1$.

The restricted regression is performed by minimizing the error sum of squares subject to the restrictions. In matrix terms, the Lagrangian for this problem is

\[
L = (y - X\beta)'(y - X\beta) + 2\lambda (R\beta - q)
\]

The Lagrangian is minimized by the restricted estimator $\beta^*$, and it can be shown that

\[
\beta^* = \hat{\beta} - (X'X)^{-1}R'\lambda
\]

where $\hat{\beta}$ is the unrestricted estimator.

Because $R\beta^* = q$, you can solve for $\lambda$ to obtain the Lagrange multipliers

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

The standard errors of the Lagrange multipliers are the square roots of the diagonal elements of the variance matrix

\[
\text{Var}(\lambda^*) = \hat{\sigma}_e^2 \left[R(X'X)^{-1}R\right]^{-1}
\]

where $\hat{\sigma}_e^2$ is the mean square error (MSE) under the null hypothesis. A significant Lagrange multiplier indicates a restriction that is binding.
Linear Hypothesis Testing

Consider a linear hypothesis of the form $R\beta = q$, where $R$ is $J \times K$ and $q$ is $J \times 1$. The Wald test statistic is

$$\chi^2_W = (R\hat{\beta} - q)' \left( R\hat{V}R' \right)^{-1} (R\hat{\beta} - q)$$

where $\hat{V}$ is the estimated variance of $\hat{\beta}$.

In simple linear models, the Wald test statistic is equal to the $F$ test statistic

$$F = \frac{(\text{SSE}_r - \text{SSE}_u)/J}{\text{SSE}_u/\text{df}_e}$$

where $\text{SSE}_r$ is the restricted error sum of squares, $\text{SSE}_u$ is the unrestricted error sum of squares, and $\text{df}_e$ is the unrestricted error degrees of freedom.

The $F$ statistic represents a more direct comparison of the restricted model to the unrestricted model. Comparing error sums of squares is appealing in complex models for which restrictions are applied not only during the final regression but also during intermediate calculations.

The likelihood ratio (LR) test and the Lagrange multiplier (LM) test are derived from the $F$ statistic. The LR test statistic is

$$\chi^2_{LR} = M \ln \left[ 1 + \frac{JF}{M-K} \right]$$

The LM test statistic is

$$\chi^2_{LM} = M \left[ \frac{JF}{M-K+JF} \right]$$

The distribution of these test statistics is $\chi^2$ with $J$ degrees of freedom. The three tests are asymptotically equivalent, but they possess different small-sample properties. For more information, see Greene (2000, p. 392) and Davidson and MacKinnon (1993, pp. 456–458).

Only the Wald is changed when a heteroscedasticity-corrected covariance matrix estimator (HCCME) is selected. The LR and LM tests are unchanged.

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 the HCCME= is specified, these variances are the same. For the two-way models, the variance and the covariances for the intercept are not corrected.\(^1\)

\(^1\)The dummy variables are removed by the within transformations, so their variances and covariances cannot be calculated the same way as the other regressors. They are recovered by the formulas in the sections “One-Way Fixed-Effects Model (FIXONE and FIXONETIME Options)” on page 1840 and “Two-Way Fixed-Effects Model (FIXTWO Option)” on page 1841. The formulas assume homoscedasticity, so they do not apply when HCCME is used. 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.
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, heteroscedasticity-corrected covariance matrix estimation (HCCME) is applied to the transformed data (\( \tilde{y} \) and \( \tilde{X} \)). In other words, first the random or fixed effects are removed through transforming the data,\(^2\) and then the heteroscedasticity (also autocorrelation with the HAC option) is corrected 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, and 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
\[ \text{Var} \left( \beta - \hat{\beta} \right) = (X'X)^{-1}X'\Omega(X'X)^{-1} \]

You can ameliorate the effect of structure in the covariance matrix by using generalized least squares (GLS), provided that \( \Omega^{-1} \) can be calculated. Using \( \Omega^{-1} \), you premultiply both sides of the regression equation,
\[ L^{-1}y = L^{-1}X\epsilon + L^{-1}\epsilon \]

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

The resulting GLS \( \beta \) is
\[ \tilde{\beta} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \]

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

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

The difference in variance between the OLS estimator and the GLS estimator can be written as

\[(X'X)^{-1}X'\Omega X (X'X)^{-1} - (X'X)^{-1}\]

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 of estimating a consistent covariance matrix and you used the OLS \(\hat{\beta}\).

Estimation of the \(\Lambda\) 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 as follows:

- **HCCME=0:**
  \[\sigma^2 X'X\]
  This is the simple OLS estimator. If you do not specify the HCCME= option, PROC PANEL defaults to this estimator.

- **HCCME=0:**
  \[
  \sum_{i=1}^{N} \sum_{t=1}^{T_i} \hat{e}_{it}^2 x_{it} x_{it}'
  + \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{s=1}^{t-1} \hat{e}_{is} \hat{e}_{is} \left( x_{is} x_{is}' + x_{is} x_{is}' \right)
  \]
  Here \(N\) is the number of cross sections and \(T_i\) is the number of observations in the \(i\)th cross section. The \(x_{it}'\) is from the \(t\)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 \(\Lambda\) is

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

  The formula is the same as the robust variance matrix estimator in Wooldridge (2002, p. 152), and it is derived under the assumptions of section 7.3.2 of Wooldridge (2002).

- **HCCME=1:**
  \[
  \frac{M}{M-K} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \hat{e}_{it}^2 x_{it} x_{it}'
  \]
  Here \(M\) is the total number of observations, \(\sum_{j=1}^{N} T_j\), and \(K\) is the number of parameters. If the CLUSTER option is specified, the estimator becomes

  \[
  \frac{M}{M-K} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \hat{e}_{it}^2 x_{it} x_{it}'
  + \frac{M}{M-K} \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{s=1}^{t-1} \hat{e}_{is} \hat{e}_{is} \left( x_{is} x_{is}' + x_{is} x_{is}' \right)
  \]
The formula is similar to the robust variance matrix estimator in Wooldridge (2002, p. 152) with the heteroscedasticity 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 \((\sum_{j=1}^{i-1} T_j + t)\)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. If the CLUSTER option is specified, the estimator becomes
  \[
  \sum_{i=1}^{N} \sum_{t=1}^{T_i} \frac{\hat{e}_{it}^2}{(1 - \hat{h}_{it})^2} x_{it} x_{it}' + 2 \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{s=1}^{t-1} \frac{\hat{e}_{it}}{1 - \hat{h}_{it}} \frac{\hat{e}_{is}}{1 - \hat{h}_{is}} \left( x_{it} x_{is}' + x_{is} x_{it}' \right)
  \]
  The formula is similar to the robust variance matrix estimator in Wooldridge (2002, p. 152) with the heteroscedasticity adjustment.

- **HCCME=3:**
  \[
  \sum_{i=1}^{N} \sum_{t=1}^{T_i} \frac{\hat{e}_{it}^2}{1 - \hat{h}_{it}} x_{it} x_{it}'
  \]
  If the CLUSTER option is specified, the estimator becomes
  \[
  \sum_{i=1}^{N} \sum_{t=1}^{T_i} \frac{\hat{e}_{it}^2}{(1 - \hat{h}_{it})^2} x_{it} x_{it}' + 2 \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{s=1}^{t-1} \frac{\hat{e}_{it}}{1 - \hat{h}_{it}} \frac{\hat{e}_{is}}{1 - \hat{h}_{is}} \left( x_{it} x_{is}' + x_{is} x_{it}' \right)
  \]
  The formula is similar to the robust variance matrix estimator in Wooldridge (2002, p. 152) with the heteroscedasticity 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 cross section, you need to take only the average of those \( N \) estimators. The details of the estimation follow. First, you arrange the data such that the first cross section occupies the first \( T_i \) observations. Then, you treat the cross sections as separate regressions with the form
  \[
  y_i = \alpha_i + X_{is}\tilde{\beta} + \epsilon_i
  \]
  where the parameter estimates \( \tilde{\beta} \) 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 \( i \)th cross section’s \( X'\Omega X \) matrix (where the \( s \) subscript indicates that no constant column has been suppressed to avoid confusion) is \( X_{is}'\Omega X_{is} \). 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, specifying HCCME=4 returns the HCCME=0 result because 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 an HCCME= option value of 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, HCCME=1 is preferred when the calculation of the hat matrix is too tedious.

All HCCMEs have well-defined asymptotic properties. The small-sample properties are not well known, and care must exercised when sample sizes are small.

The HCCME 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

$$
\Lambda_{\text{HAC}} = a \sum_{i=1}^{N} \sum_{t=1}^{T_i} \hat{e}_{it}^2 x_{it} x_{it}^\prime + a \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{s=1}^{T_i-1} k \left( \frac{s-t}{b} \right) \hat{e}_{is} \hat{e}_{is} \left( x_{it} x_{is}^\prime + x_{is} x_{it}^\prime \right)
$$

where $k(.)$ is the real-valued kernel function, $b$ is the bandwidth parameter, and $a$ is the adjustment factor of small-sample degrees of freedom (that is, $a = 1$ if the ADJUSTDF option is not specified and otherwise $a = NT/(NT - k)$, where $k$ is the number of parameters including dummy variables). The types of kernel functions are listed in Table 26.4.

<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>$k(x) = \frac{25}{12\pi^2x^2} \left( \frac{\sin(6\pi x/5)}{6\pi x/5} - \cos(6\pi x/5) \right)$</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 you specify the BANDWIDTH=ANDREWS option, the bandwidth parameter is estimated as shown in Table 26.5.

---

3Specifying HCCME=0 with the CLUSTER option sets $k(.) = 1$. 
Let \( \{g_{ait}\} \) denote each series in \( \{g_{it} = \hat{\epsilon}_{it} \hat{x}_{it}\} \), 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 terms \( \alpha(1) \) and \( \alpha(2) \) are estimated by the formulas

\[
\alpha(1) = \frac{\sum_{a=1}^{k} \frac{4 \rho_a^2 \sigma_a^4}{(1-\rho_a)^2}}{\sum_{a=1}^{k} \frac{\sigma_a^4}{(1-\rho_a)^4}} \quad \alpha(2) = \frac{\sum_{a=1}^{k} \frac{4 \rho_a^2 \sigma_a^4}{(1-\rho_a)^8}}{\sum_{a=1}^{k} \frac{\sigma_a^4}{(1-\rho_a)^4}}
\]

When you specify BANDWIDTH=NEWEYWEST94, according to Newey and West (1994) the bandwidth parameter is estimated as shown in Table 26.6.

### Table 26.6 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>

The terms \( s_0 \) and \( s_1 \) are estimated by the formulas

\[
s_0 = \sigma_0 + 2 \sum_{j=1}^{n} \sigma_j \quad s_1 = 2 \sum_{j=1}^{n} j \sigma_j
\]

where \( n \) is the lag selection parameter and is determined by kernels, as listed in Table 26.7.

### Table 26.7 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 26.7 is specified by the $C=$ option; by default, $C=12$.

The $\sigma_j$ are estimated by the equation

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

where $g_{at}$ is the same as in the Andrews method and $i$ is 1 if the NOINT option is specified in the MODEL statement, and 2 otherwise.

When you specify BANDWIDTH=SAMPLESIZE, the bandwidth parameter is estimated by the equation

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

where $T$ is the sample size; $\lfloor x \rfloor$ is the largest integer less than or equal to $x$; and $\gamma$, $r$, and $c$ are values specified by the 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

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

**R-Square**

The R-square statistic is the proportion of variability in the dependent variable that is attributed to the independent variables. Because of the transformations that are used prior to fitting the final regression model, the conventional R-square measure is not appropriate for most models that the PANEL procedure supports. In random-effects models that use a GLS transform, PROC PANEL calculates the modified R-square statistic proposed by Buse (1973),

$$
R^2 = 1 - \frac{SSE}{y'D\hat{\Omega}^{-1}Dy}
$$

where $SSE$ is the error sum of squares from the final model fit, $\hat{\Omega}^{-1/2}$ represents the GLS transform, and $D = I_M - a^{-1} J_M \hat{\Omega}^{-1}$, for $a = j_M \hat{\Omega}^{-1} j_M$.

In GLS models that do not have an intercept, the alternate R-square measure, which is attributed to Theil (1961), is calculated as follows:

$$
R^2 = 1 - \frac{SSE}{y'\hat{\Omega}^{-1}y}
$$

In fixed-effects models, the R-square measure is

$$
R^2 = 1 - \frac{SSE}{y'_w y_w}
$$
where $y_w$ is the within-transformed dependent variable.

In the case of pooled OLS estimation, all three of the R-square formulas reduce to the usual R-square statistic for linear models.

---

**F Test for No Fixed Effects**

When you fit a fixed-effects model, you obtain an $F$ test for no fixed effects as part of the output. The null hypothesis of that test is that all fixed effects are jointly 0; it is obtained by comparing fixed-effects estimates to those from pooled regression. The $F$ statistic is

$$F = \frac{(SSE_r - SSE_u)/df_1}{SSE_u/df_2} \sim F(df_1, df_2)$$

where $SSE_r$ is the error sum of squares from the restricted model (pooled regression) and $SSE_u$ is the error sum of squares from the unrestricted fixed-effects model.

The numerator degrees of freedom, $df_1$, equals $N - 1$ for one-way models and $(N - 1) + (T - 1)$ for two-way models. The denominator degrees of freedom, $df_2$, is equal to the error degrees of freedom from the fixed-effects estimation. If you specify the NOINT option, add 1 to $df_1$ to account for the added restriction to the pooled regression.

---

**Tests for Random Effects**

**Hausman Test**

For models that include random effects, the PANEL procedure outputs the results of the Hausman (1978) specification test. This test was also proposed by Wu (1973) and further extended in Hausman and Taylor (1982).

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

$$m = (\hat{\beta}_c - \hat{\beta}_e)'(\hat{\Sigma}_c - \hat{\Sigma}_e)^{-1}(\hat{\beta}_c - \hat{\beta}_e)$$

where $\hat{\Sigma}_c$ and $\hat{\Sigma}_e$ are estimates of the asymptotic covariance matrices of $\hat{\beta}_c$ and $\hat{\beta}_e$. The statistic $m$ follows a $\chi^2$ distribution with $k$ degrees of freedom, where $k$ is the rank of $(\hat{\Sigma}_c - \hat{\Sigma}_e)^{-1}$. This rank is normally equal to the dimension of $\hat{\beta}_c - \hat{\beta}_e$, but it is reduced when regressors that are constant within cross sections are dropped from the fixed-effects model.

The null hypothesis is that the effects are independent of the regressors. Under the null hypothesis, the fixed-effects estimator is consistent but inefficient, whereas the random-effects estimator is both consistent and efficient. Failure to reject the null hypothesis favors the random-effects specification.

**Breusch and Pagan Tests**

Breusch and Pagan (1980) developed 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

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

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

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

is distributed as a \( \chi^2 \) statistic with two degrees of freedom.

Because a two-way model generalizes a one-way model, failure to reject the null hypothesis of no random effects with BP2 usually implies a failure reject with BP as well. For both the BP and BP2 tests, the residuals are obtained from a pooled regression. There is very little extra cost in selecting both the BP and BP2 tests. Notice that in the case of only groupwise heteroscedasticity, the BP2 test approximates the BP test. 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 option is FIXONE or FIXONETIME. When you specify the FIXONE option, the BP option requests a test for cross-sectional random effects. When you specify the FIXONETIME option, the BP option requests a test for time random effects.

Although the Hausman test is automatically provided, you can request the Breusch-Pagan tests via the BP and BP2 options in the MODEL statement.

For more information about the Breusch and Pagan tests, see Baltagi (2013, sec. 4.2).

### Tests of Poolability

You can obtain tests for poolability across cross sections by specifying the POOLTEST option in the MODEL statement. The null hypothesis of poolability assumes homogeneous slope coefficients.

#### 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. Let \( k \) be the number of regressors except the constant. The degrees of freedom for the unrestricted model is \( df_u = M - N(k + 1) \). If the constant is restricted to be the same, the degrees of
Tests for Cross-Sectional Dependence

freedom for the restricted model is \( df_r = M - k - 1 \) and the number of restrictions is \( q = (N - 1)(k + 1) \).
If the restricted model is the fixed one-way model, the degrees of freedom is \( df_r = M - k - N \) 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, \( qF \to \chi^2(q) \). The test is the same as the Chow test (Chow 1960) extended to \( N \) linear regressions.

Likelihood Ratio (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
\[
LR = -2 \log \left( \frac{1}{(qF/df_u)^{M/2}} \right).
\]
Because \( LR = qF + O(n^{-1}) \), under the null hypothesis LR is asymptotically distributed as chi-square with \( q \) degrees of freedom.

Tests for Cross-Sectional Dependence

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 \( \hat{\rho}_{ij} \),
\[
\hat{\rho}_{ij} = \frac{\sum_{t=T_{ij}}^{T_{ij}} e_{it}e_{jt}}{\sqrt{\sum_{t=T_{ij}}^{T_{ij}} e_{it}^2 \sqrt{\sum_{t=T_{ij}}^{T_{ij}} e_{jt}^2}}}
\]
where \( T_{ij} \) and \( T_{ij} \) 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 \( T_{ij} = T \). Let \( Ti_j \) denote the number of overlapped time periods (\( T_{ij} = T_{ij} - T_{ij} + 1 \)). 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 \( Ti_j \to \infty \), \( BP \to \chi^2(N(N-1)/2) \). 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:
\[
BPs = \sqrt{\frac{1}{N(N-1)}} \sum_{i=1}^{N} \sum_{j=i+1}^{N} \left( T_{ij} \hat{\rho}_{ij}^2 - 1 \right)
\]
Under the null hypothesis, \( BPs \to \mathcal{N}(0,1) \) 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 $T_{ij} \to \infty$, $\sqrt{T_{ij} \hat{\rho}_{ij}} \to N(0,1)$. Therefore, for $N$ and $T_{ij}$ tending to infinity in any order, $CD \to N(0,1)$.

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 be applied only if the cross-sectional units can be given an ordering that remains immutable over time. Under the alternative hypothesis of a $p$th-order local dependence, the CD statistic can be generalized to a local CD test, CDp,

$$
CD_p = \sqrt{\frac{2}{p(2N-p-1)}} \left( \sum_{s=1}^{p} \sum_{i=s+1}^{N} \sqrt{T_{i,i-s} \hat{\rho}_{i,i-s}} \right)
$$

where $p = 1, \ldots, N-1$. When $p = N-1$, CDp reduces to the original CD test. Under the null hypothesis of zero cross-sectional dependence, the CDp statistic is centered at zero for fixed $N$ and $T_{i,i-s} \to k+1$, and $CD_p \to N(0,1)$ as $N \to \infty$ and $T_{i,i+s} \to \infty$.

**Panel Data Unit Root Tests**

Unit roots are a big concern in dynamic processes as they have important implications for the stationary of a process and hence estimation. Proceeding with regular estimation techniques ignoring the presence of units roots can lead to spurious regressions and hence produce nonsensical results. Therefore detecting unit roots to be able to analyze stationary processes is of vital concern for dynamic processes. One of the most widely used tests in the time series literature is the augmented Dickey-Fuller (ADF) test. This section introduces and briefly reviews the background information on the tests developed for dynamic panel data, which in most cases turn out to be enhancements of the ADF test.

**Levin, Lin, and Chu (2002)**

Levin, Lin, and Chu (2002) propose a panel unit root test for the null hypothesis of unit root against a homogeneous stationary hypothesis. The model is specified as

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

The panel unit root test evaluates the null hypothesis of $H_0 : \delta = 0$, for all $i$, against the alternative hypothesis $H_1 : \delta < 0$ for all $i$. Three models are considered: (1) $d_{1t} = \phi$ (the empty set) with no individual effects, (2) $d_{2t} = \{1\}$ in which the series $y_{it}$ has an individual-specific mean but no
time trend, and (3) $d_{3t} = \{1, t\}$ in which the series $y_{it}$ has an individual-specific mean and linear and individual-specific time trend. The lag order $p_i$ is unknown and is allowed to vary across individuals. It can be selected by the methods that are described in the section “Lag Order Selection in the ADF Regression” on page 1878. The selected lag order is denoted as $\hat{p}_i$. The necessary condition for the test is for $\sqrt{N/T} \to 0$. An important assumption is that the errors, $\varepsilon_{it}$, are assumed to be $i.i.d.(0, \sigma_{\varepsilon_{it}}^2)$. In other words, cross-sectional independence is assumed. The test is implemented in the following three steps:

**Step 1** The ADF regressions are implemented for each individual $i$, and then the orthogonalized residuals are generated and normalized. That is, the following model is estimated:

$$\Delta y_{it} = \delta_i y_{it-1} + \sum_{L=1}^{\hat{p}_i} \theta_{iL} \Delta y_{it-L} + \alpha_{mi} d_{mt} + \varepsilon_{it} \quad m = 1, 2, 3$$

Then, two orthogonalized residuals are generated by the following two auxiliary regressions:

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

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

The residuals are then saved as $\hat{e}_{it}$ and $\hat{v}_{it-1}$, respectively, then normalized using the regression standard error from the ADF regression in order to remove heteroscedasticity. Let $\hat{\sigma}_{\varepsilon_i}$ denote the standard error from each of the previous ADF regressions, where $\hat{\sigma}_{\varepsilon_i}^2 = \frac{\hat{\varepsilon}_{it} - \hat{\delta}_i \hat{\varepsilon}_{it-1}}{(T - p_i - 1)}$. The normalized residuals are then:

$$\hat{e}_{it} = \frac{\hat{e}_{it}}{\hat{\sigma}_{\varepsilon_i}}, \quad \hat{v}_{it-1} = \frac{\hat{v}_{it-1}}{\hat{\sigma}_{\varepsilon_i}}$$

**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 $s_{\gamma_{\varepsilon_i}}$, 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 1879. Then the ratios are estimated by $\hat{s}_i = \frac{\hat{s}_{\gamma_{\varepsilon_i}}}{\hat{\sigma}_{\varepsilon_i}}$. Let the average standard deviation ratio be $S_N = (1/N) \sum_{i=1}^{N} s_i$, and let its estimator be $\hat{S}_N = \frac{1}{N} \sum_{i=1}^{N} \hat{s}_i$. As the authors note in their paper, use of the long run variance based on first-differences results in lower bias in finite samples.

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

$$\hat{e}_{it} = \delta \hat{v}_{it-1} + \hat{e}_{it}$$

The total number of observations is $N \bar{T}$, with $\bar{T} = \frac{\sum_{i=1}^{N} \hat{p}_i}{N}$. The standard $t$ statistic for testing $H_0 : \delta = 0$ is $t_{\delta} = \hat{\delta}/\hat{\sigma}_{\delta}$, with OLS estimator $\hat{\delta}$ and standard deviation $\hat{\sigma}_{\delta}$. 


\[ \hat{\delta} = \frac{\sum_{i=1}^{N} \sum_{t=2+\hat{p}_i}^{T} \hat{\epsilon}_{it} \hat{\nu}_{it-1}}{\sum_{i=1}^{N} \sum_{t=2+\hat{p}_i}^{T} \hat{\nu}^2_{it-1}} \]

\[ \hat{\delta} = \hat{\delta} \left[ \sum_{i=1}^{N} \sum_{t=2+\hat{p}_i}^{T} \hat{\nu}^2_{it-1} \right]^{-\frac{1}{2}} \]

Where \( \hat{\delta} \) be the root mean square error from the step 3 regression

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

However, the standard t statistic diverges to negative infinity for models (2) and (3). Levin, Lin, and Chu (2002) therefore propose the following adjusted t statistic:

\[ t^*_{\delta} = \frac{t_{\delta} - N \bar{T} \hat{\delta} \hat{\nu}^2 - 2 \hat{\delta} \hat{\nu}^2 \mu^*_{mT}}{\sigma^*_{mT}} \]

The mean and standard deviation adjustments \((\mu^*_{mT}, \sigma^*_{mT})\) depend on the time series dimension \( \bar{T} \) and model specification \( m \), which can be found in Table 2 of Levin, Lin, and Chu (2002). The adjusted t statistic converges to the standard normal distribution. Therefore, the standard normal critical values are used in hypothesis testing.

**Lag Order Selection in the ADF Regression**

The methods for selecting the individual lag orders in the ADF regressions can be divided into two categories: selection based on information criteria and selection via sequential testing.

**Lag Selection Based on Information Criteria** In this method, the following information criteria can be applied to lag order selection: AIC, SBC, HQIC (HQC), and MAIC. As with other model selection applications, the lag order is selected from 0 to the maximum \( p_{max} \) to minimize the objective function, plus a penalty term, which is a function of the number of parameters in the regression. Let \( k \) be the number of parameters and \( T_o \) be the number of effective observations. For regression models, the objective function is \( T_o \log(SSR/T_o) \), where SSR is the sum of squared residuals. For AIC, the penalty term equals \( 2k \). For SBC, this term is \( k \log T_o \). For HQIC, it is \( 2c \log \log(T_o) \) with \( c \) being a constant greater than 1.\(^4\) For MAIC, the penalty term equals \( 2(\tau_T(k) + k) \), where

\[ \tau_T(k) = (SSR/T_o)^{-1} \sum_{t=p_{max}+2}^{T} y^2_{t-1} \]

and \( \hat{\delta} \) is the estimated coefficient of the lagged dependent variable \( y_{t-1} \) in the ADF regression.

\(^4\)In practice \( c \) is set to 1, following the literature (Hannan and Quinn 1979; Hall 1994).
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 \( p_{\text{max}} \), the \( t \) test for the largest lag order in \( \hat{\theta}_i \) is performed to determine whether a smaller lag order is preferred. Specifically, when the null of \( \hat{\theta}_iL = 0 \) is not rejected given the significance level (5%), a smaller lag order is preferred. This procedure continues until a statistically significant lag order is reached. On the other hand, the SG modeling strategy starts with lag order 0 and moves toward the maximum lag order \( p_{\text{max}} \).

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 \( \bar{K} \) and kernel weights \( w_{\bar{K}L} \), the formula is

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

To achieve consistency, the lag truncation parameter must satisfy \( \bar{K}/T \to 0 \) and \( \bar{K} \to \infty \) as \( T \to \infty \). Levin, Lin, and Chu (2002) suggest \( \bar{K} = \left[ 3.21T^{1/3} \right] \). The weights \( w_{\bar{K}L} \) depend on the kernel function. Andrews (1991) proposes data-driven bandwidth (lag truncation parameter + 1 if integer-valued) selection procedures to minimize the asymptotic mean squared error (MSE) criterion. For more information about the kernel functions and Andrews (1991) data-driven bandwidth selection procedure, see the section “Heteroscedasticity- and Autocorrelation-Consistent Covariance Matrices” on page 1870. Because Levin, Lin, and Chu (2002) truncate the bandwidth as an integer, when LLCBAND is specified as the BANDWIDTH option, it corresponds to \( \text{BANDWIDTH} = \left[ 3.21T^{1/3} \right] + 1 \). Furthermore, kernel weights \( w_{\bar{K}L} = k(L/(\bar{K}+1)) \) with kernel function \( k() \).

For model (2), the series \( \Delta y_{it} \) is demeaned individual by individual first. Therefore, \( \Delta y_{it} \) is replaced by \( \Delta y_{it} - \overline{\Delta y_{it}} \), where \( \overline{\Delta y_{it}} \) 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 \( \overline{\Delta y_{it}} \), 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

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

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 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:

\[
\Delta y_{it} = \beta_i y_{it-1} + \sum_{j=1}^{p_i} \rho_{ij} \Delta y_{i,t-j} + \alpha_m d_m + \varepsilon_{it} \quad m = 1, 2, 3
\]

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 \( t-bar_{NT} \), 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

\[
t-bar_{NT} = N^{-1} \sum_{i=1}^{N} t_i (p_i, \rho_i)
\]

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

\[
\eta_i = \frac{1}{2} \{ [W_i(1)]^2 - 1 \} - W_i(1) \int_0^1 W_i(u) du - \frac{1}{2} \int_0^1 [W_i(u)]^2 du - \frac{1}{4} \int_0^1 W_i(u) du \]  \( \Rightarrow N(0, 1) \)

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

\[
Z_{t-bar}(p, \rho) = \frac{\sqrt{N} \{ t-bar_{NT} - E(\eta) \}}{\sqrt{\text{Var}(\eta)}} \Rightarrow N(0, 1)
\]
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 \( \bar{t} \)-statistic by means and variances of \( t_iT(p_i, 0) \) under the null hypothesis \( \beta_i = 0 \). The alternative standardized \( \bar{t} \)-statistic is

\[
W_{\bar{t}-bar}(p, \rho) = \sqrt{N} \left\{ \bar{t}_{-bar NT} - N^{-1} \sum_{i=1}^{N} E[t_iT(p_i, 0) | \beta_i = 0] \right\} \overset{\text{N}(0, 1)}{\longrightarrow} \Rightarrow \]

Im, Pesaran, and Shin (2003) simulate the values of \( E[t_iT(p_i, 0) | \beta_i = 0] \) and \( \text{Var}[t_iT(p_i, 0) | \beta_i = 0] \) 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). For more information, see the section “Lag Order Selection in the ADF Regression” on page 1878.

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 \( \bar{\hat{\beta}}_i \) as the estimated standard error from the restricted regression \( \bar{\beta}_i = 0 \),

\[
\bar{\hat{\beta}}_i = \frac{\hat{\beta}_{iT} \left( y_{i,-1}^T M_{t_i} y_{i,-1} \right)^{1/2}}{\bar{\hat{\sigma}}_iT}
\]

where \( \hat{\beta}_{iT} \) is the OLS estimator of \( \beta_i \) (unrestricted model), \( \tau_T = (1, 1, \ldots, 1)^T \), \( M_t = I_T - \tau_T (\tau_T^T \tau_T)^{-1} \tau_T^T \), and \( y_{i,-1} = (y_{i0}, y_{i1}, \ldots, y_{iT-1})^T \). Under the null hypothesis, the standardized \( \bar{t} \)-statistic converges to a standard normal variate,

\[
Z_{\bar{t}-bar} = \frac{\sqrt{N} \left\{ \bar{\hat{\beta}}_i - E \left( \bar{\hat{\beta}}_i \right) \right\}}{\sqrt{\text{Var} \left( \bar{\hat{\beta}}_i \right)}} \overset{\text{N}(0, 1)}{\longrightarrow} \Rightarrow
\]

where \( E \left( \bar{\hat{\beta}}_i \right) \) and \( \text{Var} \left( \bar{\hat{\beta}}_i \right) \) are the mean and variance of \( \bar{\hat{\beta}}_i \), 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 \( Z_{\bar{t}-bar} \) 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 \( \bar{t}_{-bar NT} \) are simulated.

Similar as in section “Levin, Lin, and Chu (2002)” on page 1876, it is possible to relax this assumption of cross-sectional independence and allow for a limited degree of dependence via time-specific aggregate effects. Two more models (model 4 and model 5) with time fixed effects are considered. For more information, see the section “Cross-Sectional Dependence via Time-Specific Aggregate Effects” on page 1879.

**Combination Tests**

Combining the observed significance levels (\( p \)-values) from \( N \) independent tests of the unit root null hypothesis was proposed by Maddala and Wu (1999); Choi (2001). Suppose \( G_i \) is the test statistic to test the unit root null hypothesis for individual \( i = 1, \ldots, N \), and \( F(\cdot) \) is the cdf (cumulative distribution function) of the asymptotic distribution as \( T \to \infty \). Then the asymptotic \( p \)-value is defined as

\[
p_i = F \left( G_i \right)
\]

There are different ways to combine these \( p \)-values. The first one is the inverse chi-square test (Fisher 1932); this test is referred to as \( P \) test in Choi (2001) and \( \lambda \) in Maddala and Wu (1999):

\[
P = -2 \sum_{i=1}^{N} \ln \left( p_i \right)
\]
When the test statistics \( \{G_i\}_{i=1}^N \) are continuous, \( \{p_i\}_{i=1}^N \) are independent uniform (0, 1) variables. Therefore, \( P \Rightarrow \chi^2_{2N} \) as \( T \to \infty \) and \( N \) fixed. But as \( N \to \infty \), \( P \) diverges to infinity in probability. Therefore, it is not applicable for large \( N \). To derive a nondegenerate limiting distribution, the \( P \) test (Fisher test with \( N \to \infty \)) should be modified to

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

Under the null as \( T_i \to \infty \), and then \( N \to \infty \), \( P_m \Rightarrow N(0,1) \).

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 \( T_i \to \infty \), \( Z \Rightarrow N(0,1) \) 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 \( k = 3(5N + 4) / (\pi^2 N (5N + 2)) \). When \( T_i \to \infty \) and \( N \) is fixed, \( L^* \Rightarrow t_{5N+4} \). In other words, the limiting distribution is the \( t \) distribution with degree of freedom \( 5N + 4 \). The sequential limit is \( L^* \Rightarrow N(0,1) \) 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 1876, it is possible to relax this assumption of cross-sectional independence and allow for a limited degree of dependence via time-specific aggregate effects. Two more models (model 4 and model 5) with time fixed effects are considered. For more information, see the section “Cross-Sectional Dependence via Time-Specific Aggregate Effects” on page 1879.

**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

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

\[5\] The time series length \( T \) is subindexed by \( i = 1, \ldots, N \) because the panel can be unbalanced.

\[6\] Choi (2001) also points out that the joint limit result where \( N \) and \( \{T_i\}_{i=1}^N \) go to infinity simultaneously is the same as the sequential limit, but it requires more moment conditions.
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:

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

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

\[
\hat{\sigma}_i^2 = \frac{T}{T - 2} \left( \Delta y_{it} - \sum_{t=2}^{T} \Delta y_{it} / (T - 1) \right)^2 / (T - 2)
\]

The UB statistic has a standard normal limiting distribution as \( T \to \infty \) followed by \( N \to \infty \) sequentially.

To account for the short-run dynamics, Breitung and Das (2005) suggest applying the test to the prewhitened series, \( \hat{y}_{it} \). For model (1) and model (2) (constant-only case), they suggested the same method as in step 1 of Levin, Lin, and Chu (2002).\(^7\) 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 (\( \mu_i = 0, \beta_i = 0 \)):

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

where \( \hat{p}_i \) is a consistent estimator of the true lag order \( p_i \) and can be estimated by the procedures listed in the section “Lag Order Selection in the ADF Regression” on page 1878. For LLC and IPS tests, the lag orders are selected by running the ADF regressions. But for Breitung and his coauthors’ tests, the restricted ADF regressions are used to be consistent with the prewhitening method. Let \( \left( \hat{\mu}_i, \hat{\theta}_{iL} \right) \) be the estimated coefficients.\(^8\) The prewhitened series can be obtained by

\[
\Delta \hat{y}_{it} = \Delta y_{it} - \sum_{L=1}^{\hat{p}_i} \hat{\theta}_{iL} \Delta y_{i,t-L}
\]

\(^7\)For more information, see the section “Levin, Lin, and Chu (2002)” on page 1876. The only difference is the standard error estimate \( \hat{\sigma}_i^2 \). Breitung suggests using \( T - p_i - 2 \) instead of \( T - p_i - 1 \) as in LLC to normalize the standard error.

\(^8\)Breitung (2000) suggests the approach in step 1 of Levin, Lin, and Chu (2002), while Breitung and Das (2005) suggest the prewhitening method as described above. In Breitung’s code, to be consistent with the papers, different approaches are adopted for model (2) and (3). Meanwhile, for the order of variable transformation and prewhitening, in model (2), the initial values are deducted (variable transformation) first, and then the prewhitening was applied. For model (3), the order is reversed. The series is prewhitened and then transformed to remove the mean and linear time trend.
and
\[ \hat{y}_{it} = y_{it} - \sum_{L=1}^{\hat{\theta}_L} \hat{\theta}_L y_{it-L} \]
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 \( N \to \infty \),
\[ t_{OLS} = \frac{\sum_{t=1}^{T} \sum_{i=1}^{N} y_{it-1} \Delta y_{it}}{\hat{\sigma} \sqrt{\sum_{t=1}^{T} \sum_{i=1}^{N} y_{it-1}^2}} \to N(0,1) \]
where \( \hat{\sigma}^2 = \frac{\sum_{t=2}^{T} (\Delta y_{it} - \delta \hat{y}_{i,t-1})^2}{N(T-1)} \) 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 \( \Omega = E(v_{1t}v_{it}') \) be a positive definite matrix with eigenvalues \( \lambda_1 \geq \cdots \geq \lambda_N \). Let \( y_t = (y_{1t}, \ldots, y_{Nt})' \) and \( \Delta y_t = (\Delta y_{1t}, \ldots, \Delta y_{Nt})' \). The model can be written as a SUR-type system of equations,
\[ \Delta y_t = \delta y_{t-1} + v_t \]
The unknown covariance matrix \( \Omega \) can be estimated by its sample counterpart,
\[ \hat{\Omega} = \sum_{t=2}^{T} (\Delta y_{t-1} - \hat{\delta} y_{t-1}) (\Delta y_{t-1} - \hat{\delta} y_{t-1})' / (T-1) \]
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 \( t_\Omega = \lim_{N \to \infty} \text{tr} (\Omega^2 / N) / (\text{tr} \Omega^2 / N)^2 \). The variance \( t_\Omega \) can be consistently estimated by
\[ \hat{t}_\delta = \left( \sum_{t=2}^{T} y_{t-1}' \hat{\Omega} y_{t-1} \right) / \left( \sum_{t=2}^{T} y_{t-1}' \hat{\Omega} y_{t-1} \right)^2 \]
Thus the robust \( t \) statistic can be calculated as
\[ t_{rob} = \frac{\hat{\delta}}{\hat{t}_\delta} = \frac{\sum_{t=2}^{T} y_{t-1}' \Delta y_{t}}{\sqrt{\sum_{t=2}^{T} y_{t-1}' \hat{\Omega} y_{t-1}}} \to N(0,1) \]
as \( T \to \infty \) followed by \( N \to \infty \) under the null hypothesis of random walk. Since the finite sample distribution can be quite different, Breitung and Das (2005) list the 1%, 5%, and 10% critical values for different \( N \)'s.
When \( T > N \), a (feasible) GLS estimator is applied; it is asymptotically more efficient than the OLS estimator. The data are transformed by multiplying \( \hat{\Omega}^{-1/2} \) as defined before, \( \hat{\Omega}^{-1/2} y_t \). Thus the model is transformed into
\[ \Delta \hat{z}_t = \delta \hat{z}_{t-1} + e_t \]
The feasible GLS (FGLS) estimator of \( \delta \) and the corresponding \( t \) statistic are obtained by estimating the transformed model by OLS and denoted by \( \hat{\delta}_{GLS} \) and \( t_{GLS} \), respectively:
\[ t_{GLS} = \frac{\sum_{t=2}^{T} y_{t-1}' \hat{\Omega}^{-1} \Delta y_{t}}{\sqrt{\sum_{t=2}^{T} y_{t-1}' \hat{\Omega}^{-1} y_{t-1}}} \to N(0,1) \]
Similar as in section “Levin, Lin, and Chu (2002)” on page 1876, it is possible to relax this assumption of cross-sectional independence and allow for a limited degree of dependence via time-specific aggregate effects. Two more models (model 4 and model 5) with time fixed effects are considered. For more information, see the section “Cross-Sectional Dependence via Time-Specific Aggregate Effects” on page 1879.

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 stationarity, the variance of the random walk equals 0. Specifically, two models are considered:

- For model (1), the time series \( y_{it} \) is stationary around a level \( r_{i0} \),
  \[
  y_{it} = r_{it} + \epsilon_{it} \quad i = 1, \ldots, N, \quad t = 1, \ldots, T
  \]
- For model (2), \( y_{it} \) is trend stationary,
  \[
  y_{it} = r_{it} + \beta_i t + \epsilon_{it} \quad i = 1, \ldots, N, \quad t = 1, \ldots, T
  \]
  where \( r_{it} \) is the random walk component,
  \[
  r_{it} = r_{it-1} + u_{it} \quad i = 1, \ldots, N, \quad t = 1, \ldots, T
  \]
  The initial values of the random walks, \( \{r_{i0}\}_{i=1}^{N} \), 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 \( y_i' = (y_{i1}, \ldots, y_{iT}) \), \( e_i' = (\epsilon_{i1}, \ldots, \epsilon_{iT}) \) with \( \epsilon_{it} = \sum_{j=1}^{T} u_{ij} + \epsilon_{it} \), and \( X_i = (t_T, a_T) \) with \( t_T \) being a \( T \times 1 \) vector of ones, \( a_T' = (1, \ldots, T) \), and \( \beta_i' = (r_{i0}, \beta_i) \).

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

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

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

\[
LM \xrightarrow{p} E \left[ \int_0^1 V^2 (r) \, dr \right]
\]

where \( V (r) \) is a standard Brownian bridge in model (1) and a second-level Brownian bridge in model (2). Let \( W (r) \) be a standard Wiener process (Brownian motion),

\[
V (r) = \begin{cases} 
W (r) - rW (1) & \text{for model (1)} \\
W (r) + (2r - 3r^2) W (1) + 6r (r - 1) \int_0^1 W (s) \, ds & \text{for model (2)}
\end{cases}
\]
The mean and variance of the random variable $\int V^2$ can be calculated by using the characteristic functions,

$$\xi = E \left[ \int_0^1 V^2 (r) \, dr \right] = \begin{cases} \frac{1}{6} & \text{for model (1)} \\ \frac{1}{15} & \text{for model (2)} \end{cases}$$

and

$$\xi^2 = var \left[ \int_0^1 V^2 (r) \, dr \right] = \begin{cases} \frac{1}{45} & \text{for model (1)} \\ \frac{1}{6300} & \text{for model (2)} \end{cases}$$

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

$$Z = \frac{N (LM - \xi)}{\xi} \quad \implies \quad N(0, 1)$$

**Consistent Estimator of $\sigma^2_e$**

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

$$\hat{\sigma}^2_e = \sum_{i=1}^N \sum_{t=1}^T \hat{\epsilon}^2_{it} / N (T - k)$$

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

$$LM = \frac{1}{N} \sum_{i=1}^N \left( \frac{1}{T^2} \sum_{t=1}^T S_{it}^2 / \hat{\sigma}^2_{e,i} \right)$$

To allow for temporal dependence over $t$, $\sigma^2_e$ has to be replaced by the long-run variance of $\epsilon_{it}$, which is defined as $\sigma^2 = \sum_{i=1}^N \lim_{T \to \infty} T^{-1} (S_{iT}^2) / N$. 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 1879.

Similar as in section “Levin, Lin, and Chu (2002)” on page 1876, it is possible to relax this assumption of cross-sectional independence and allow for a limited degree of dependence via time-specific aggregate effects. One more models (model 3) with time fixed effects are considered. For more information, see the section “Cross-Sectional Dependence via Time-Specific Aggregate Effects” on page 1879.

**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,

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

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

$$y_{it} = \alpha_i + \varphi y_{it-1} + \nu_{it}$$
Model (3) includes heterogeneous drifts and linear time trends,
\[ y_{it} = \alpha_i + \beta_i t + \varphi y_{it-1} + v_{it} \]

Similar as in section “Levin, Lin, and Chu (2002)” on page 1876, it is possible to relax this assumption of cross-sectional independence and allow for a limited degree of dependence via time-specific aggregate effects. Two more models (model 4 and model 5) with time fixed effects are considered. For more information, see the section “Cross-Sectional Dependence via Time-Specific Aggregate Effects” on page 1879.

Let \( \hat{\phi} \) be the OLS estimator of \( \varphi \); then
\[
\hat{\phi} - 1 = \left[ \sum_{i=1}^{N} y_{i,-1} Q_T y_{i,-1} \right]^{-1} \cdot \left[ \sum_{i=1}^{N} y_{i,-1} Q_T v_i \right]
\]

where \( y_{i,-1} = (y_{i0}, \ldots, y_{iT-1}) \), \( v_i = (v_{i1}, \ldots, v_{iT}) \), and \( Q_T \) is the projection matrix. For model (1), there are no regressors other than the lagged dependent value, so \( Q_T = I_T \). 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 \( N \rightarrow \infty \)
\[
\sqrt{NT} (T - 1)/2 (\hat{\phi} - 1) \xrightarrow{y_{i0}=0,H_0} \mathcal{N} (0, 1)
\]

As \( T \rightarrow \infty \), it becomes \( T \sqrt{N} (\hat{\phi} - 1) \xrightarrow{H_0} \mathcal{N} (0, 2) \).

When the drift is absent in model (2), \( \alpha_i = 0 \), under the null hypothesis, as \( N \rightarrow \infty \)
\[
\sqrt{\frac{5N}{3}} \left( \frac{(T + 1)^3 (T - 1)}{17T^2 - 20T + 17} \right) (\hat{\phi} - 1 + \frac{3}{(T + 1)}) \xrightarrow{\alpha_i=0,H_0} \mathcal{N} (0, 1)
\]

As \( T \rightarrow \infty \), \( \left( T \sqrt{\frac{N}{5}} (\hat{\phi} - 1) + 3\sqrt{\frac{N}{5}} \right) / \sqrt{\frac{51}{51}} \xrightarrow{H_0} \mathcal{N} (0, 1) \).

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

When \( T \rightarrow \infty \), \( \left( T \sqrt{\frac{N}{2895}} (\hat{\phi} - 1) + 7.5\sqrt{\frac{N}{2895}} \right) / \sqrt{\frac{2895}{112}} \xrightarrow{H_0} \mathcal{N} (0, 1) \).
Honda (1985) and Honda (1991) UMP Test and Moulton and Randolph (1989) SLM Test

The Breusch-Pagan LM test is two-sided when the variance components are nonnegative. For a one-sided alternative hypothesis, Honda (1985) suggests a uniformly most powerful (UMP) LM test for $H_0^1: \sigma_y^2 = 0$ (no cross-sectional effects) that is based on the pooled estimator. The alternative is the one-sided $H_1^1: \sigma_y^2 > 0$. Let $\hat{u}_{it}$ be the residual from the simple pooled OLS regression and

$$d = \left( \sum_{i=1}^{N} \left[ \sum_{t=1}^{T} \hat{u}_{it}^2 \right] \right) / \left( \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2 \right).$$

Then the test statistic is defined as

$$J = \sqrt{\frac{NT}{2(T-1)}} [d - 1] \overset{H_0^1}{\rightarrow} N(0, 1)$$

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

$$S = \frac{J - E(J)}{\sqrt{\text{Var}(J)}} = \frac{d - E(d)}{\sqrt{\text{Var}(d)}} \rightarrow \mathcal{N}(0, 1)$$

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

$$\text{Tr}(DM_Z) = NT - \text{Tr} \left( J_T \sum_{i=1}^{N} \left[ Z_i \left( \sum_{j=1}^{N} Z_j'Z_j \right)^{-1} Z_i' \right] \right)$$

To test for $H_0^2: \sigma_y^2 = 0$ (no time effects), define $d2 = \left( \sum_{t=1}^{T} \left[ \sum_{i=1}^{N} \hat{u}_{it}^2 \right] \right) / \left( \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2 \right)$. Then the test statistic is modified as

$$J2 = \sqrt{\frac{NT}{2(N - 1)}} [d2 - 1] \overset{H_0^2}{\rightarrow} N(0, 1)$$

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

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

To test for $H_0^3: \sigma_y^2 = 0, \sigma_u^2 = 0$ (no cross-sectional and time effects), the test statistic is $J3 = (J + J2)/\sqrt{2}$ and $D = \sqrt{n / (T - 1)}(J_N \otimes J_T) / 2 + \sqrt{n / (N - 1)}(J_N \otimes I_T) / 2$. To standardize, define $d3 = \sqrt{n / (T - 1)}d/2 + \sqrt{n / (N - 1)}(d2)/2$,

$$S3 = \frac{J3 - E(J3)}{\sqrt{\text{Var}(J3)}} = \frac{d3 - E(d3)}{\sqrt{\text{Var}(d3)}} \rightarrow \mathcal{N}(0, 1)$$
King and Wu (1997) **LMMP Test and the SLM Test**

King and Wu (1997) derive the locally mean most powerful (LMMP) one-sided test for $H_0^1$ and $H_0^2$, which coincides with the Honda (1985) UMP test. Baltagi, Chang, and Li (1992) extend the King and Wu (1997) test for $H_3^0$ as follows:

$$KW = \frac{\sqrt{T-1}}{\sqrt{N+T-2}}J + \frac{\sqrt{N-1}}{\sqrt{N+T-2}}J2 \overset{H_0^3}{\rightarrow} N(0,1)$$

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

$$S_{kw} = \frac{KW - E(KW)}{\sqrt{\text{Var}(KW)}} = \frac{d_{kw} - E(d_{kw})}{\sqrt{\text{Var}(d_{kw})}} \overset{H_0^3}{\rightarrow} N(0,1)$$

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

If one or both variance components ($\sigma^2$ and $\alpha^2$) 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

$$GHM = \begin{cases} 
J^2 + (J2)^2 & \text{if } J > 0, J2 > 0 \\
J^2 & \text{if } J > 0, J2 \leq 0 \\
(J2)^2 & \text{if } J \leq 0, J2 > 0 \\
0 & \text{if } J \leq 0, J2 \leq 0 
\end{cases} \overset{H_0^3}{\rightarrow} \left(\frac{1}{4}\right) \chi^2(0) + \left(\frac{1}{2}\right) \chi^2(1) + \left(\frac{1}{4}\right) \chi^2(2)$$

where $\chi^2(0)$ is the unit mass at the origin.

**Tests for Serial Correlation and Cross-Sectional Effects**

The presence of cross-sectional effects causes serial correlation in the errors. Therefore, serial correlation is often tested jointly with cross-sectional effects. Joint and conditional tests for both serial correlation and cross-sectional effects have been covered extensively in the literature.

**Baltagi and Li Joint LM Test for Serial Correlation and Random Cross-Sectional Effects**

Baltagi and Li (1991) derive the LM test statistic, which jointly tests for zero first-order serial correlation and random cross-sectional effects under normality and homoscedasticity. The test statistic is independent of the form of serial correlation, so it can be used with either AR(1) or MA(1) error terms. The null hypothesis is a white noise component: $H_0^1: \sigma^2 = 0$, $\theta = 0$ for MA(1) with MA coefficient $\theta$ or $H_0^2: \alpha^2 = 0$, $\rho = 0$ 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

$$BL.91 = \frac{NT^2}{2(T-1)(T-2)} \left[A^2 - 4AB + 2TB^2\right] \overset{H_0^{1,2}}{\rightarrow} \chi^2(2)$$

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

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

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 1888, 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

$$ RS^*_\rho = \frac{NT^2 (B - A/T)^2}{(T-1)(1-2/T)} $$

Baltagi and Li (1991, 1995) derive the conventional RS test when the cross-sectional random effects is assumed to be absent:

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

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 $\sigma^2 > 0$, 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:

$$ RSO^*_\mu = \sqrt{\frac{NT}{2(T-1)(1-2/T)}} (A - 2B) \rightarrow N(0,1) $$
Baltagi and Li (1995) **LM Test for First-Order Correlation under Fixed Effects**

The two-sided LM test statistic for testing a white noise component in a fixed one-way model ($H_0^5 : \theta = 0$ or $H_0^6 : \rho = 0$, given that $\gamma_t$ are fixed effects) is

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

where $\hat{u}_{it}$ are the residuals from the fixed one-way model (FIXONE). 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 Test**

Bhargava, Franzini, and Narendranathan (1982) propose a test of serial correlation using the Durbin-Watson statistic

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

where $\hat{e}_{it}$ are the residuals from the fixed one-way model (FIXONE).

The test statistic $d_\rho$ ranges between 0 and 4, where $d_\rho = 2$ indicates no serial correlation. Values closer to 0 indicate positive serial correlation while values closer to 4 indicate negative serial correlation. A value of 0 indicates a random walk.

The PANEL procedure outputs three Durbin-Watson tests for serial correlation:

1. **White Noise vs. Positive Correlation**: $H_0 : \rho = 0$ vs. $H_1 : \rho > 0$
2. **Random Walk vs. Stationary**: $H_0 : \rho = 1$ vs. $H_1 : \rho < 1$
3. **White Noise vs. Negative Correlation**: $H_0 : \rho = 0$ vs. $H_1 : \rho < 0$

The first two tests report $d_\rho$ as the test statistic, while the third test reports $4 - d_\rho$, where values of $4 - d_\rho$ close to 0 indicate negative correlation.

In finite samples, the mechanics of the Durbin-Watson test produce an indeterminate region, a region of uncertainty as to whether to reject the null hypothesis. Because of this ambiguity, each test reports two $p$-values: The first, Pr < DWLower, treats the uncertainty region as a rejection region. The second, Pr > DWUpper, is more conservative and treats the uncertainty region as a failure-to-reject region. You can think of these two $p$-values as bounds on the exact $p$-value.
Berenblut-Webb Statistic

Bhargava, Franzini, and Narendranathan (1982) also suggest using the Berenblut-Webb statistic, which is a locally most powerful invariant test in the neighborhood of $\rho = 1$. The test statistic is

$g_\rho = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} \Delta \hat{u}_{i,t}^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{i,t}^2}$

where $\Delta \hat{u}_{i,t}$ are the residuals from the first-difference estimation. The upper and lower bounds are the same as for the Durbin-Watson statistic $d_\rho$ and produce two $p$-values, one conservative and one anti-conservative.

Testing for Random Walk Null Hypothesis

You can also use the Durbin-Watson and Berenblut-Webb statistics to test the random walk null hypothesis, with the bounds that are listed in Bhargava, Franzini, and Narendranathan (1982). For more information about these statistics, see the sections “Durbin-Watson Test” on page 1891 and “Berenblut-Webb Statistic” on page 1892. 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 $(T - 1) \times (T - 1)$ symmetric matrix that has the following elements:

$F_{tt'} = \begin{cases} (T - t') / T & \text{if } t' \geq t \\ 0 & \text{otherwise} \end{cases}$

$\left( t, t' = 1, \ldots, T - 1 \right)$

The test statistic is

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

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

In general, there must be at least one cross section that has more than one time series observation. Some estimation methods might have more stringent requirements; for example, the Amemiya-MacCurdy estimator requires data that are balanced. Some estimators require that there be more cross sections than time series values. When the data are insufficient for an estimator, check the log for error messages that provide further details.

If you are using the Parks method (by specifying the PARKS option in the MODEL statement) 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 a seemingly unrelated regression that has fewer observations than equations in the model. To avoid this problem, reduce the number of cross sections.

It is vitally important that you sort your data by cross sections and by time periods within cross sections. As PROC PANEL steps through the observations in the data, it treats any change in the value of the cross section ID variable as a new cross section, regardless of whether it has encountered that value previously. If you do not sort your data, the results might not be what you expect.
PROC PANEL is not supported for data sets that have duplicated time values within cross sections. If data with such duplication are encountered, PROC PANEL issues an error message such as the following:

“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.”

### 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. Table 26.8 lists the graph names, the plot descriptions, and the options used.

<table>
<thead>
<tr>
<th>ODS Graph Name</th>
<th>Plot Description</th>
<th>PLOTS= Options</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>RESIDHISTOGRAM, RESIDHISTOGRAM</td>
</tr>
</tbody>
</table>

### OUTPUT OUT= Data Set

PROC PANEL writes the initial data of the estimated model, predicted values, and residuals to an output data set when you specify the OUT= option in the OUTPUT statement. The OUT= data set contains the following variables:

- `_MODELL_` is a character variable that contains the label for the MODEL statement if a label is specified.
- `_METHOD_` is a character variable that identifies the estimation method.
Chapter 26: The PANEL Procedure

_MODLNO_ is the number of the model estimated.

_ACTUAL_ contains the value of the dependent variable.

_WEIGHT_ contains the weighting variable.

_CSID_ is the value of the cross section ID.

_TSID_ is the value of the time period in the dynamic model.

_regressors_ are the values of regressor variables specified in the MODEL statement.

_name_ if PRED= name1 and/or RESIDUAL= name2 options are specified, then name1 and name2 are the columns of predicted values of dependent variable and residuals of the regression, respectively.

OUTEST= Data Set

PROC PANEL writes the parameter estimates to an output data set when you specify the OUTEST= option in the PROC PANEL statement. 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 this 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. This variable is used in conjunction with the _TYPE_ variable 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. This variable is used with the _TYPE_ variable 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. This variable is included in the OUTEST= data set when a one-way or two-way random-effects model is estimated.

_VARTS_ is the variance component estimate due to time series. This 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. This variable is included in the OUTTEST= data set when a one-way or two-way random-effects model is estimated.

_A_1 is the first-order autoregressive parameter estimate. This variable is included in the OUTTEST= 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 OUTTEST= 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**

If you specify the FIXONE, FIXONETIME, FDONE, FDONETIME, or RANONE option and the OUTTRANS= option, the transformed dependent variable and independent variables are written to a SAS data set; other variables in the input data set are copied unchanged.

Suppose your data set contains the variables y, x1, x2, x3, and z2. The following statements create a SAS data set that contains the transformed data:

```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 by their deviations from the cross-sectional means. Furthermore, the following new variables are created:

_MODEL_ is the model’s label (if it exists).

_METHOD_ is the model’s transformation type. This variable reflects the estimation method and, in the case of random effects, the variance-component method.

**Printed Output**

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

- a model description, which gives the estimation method used, the model statement label if specified, the number of cross sections and number of observations in each cross section, and the order of the 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, these include 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 that it creates. You can use this name to refer to the table when using the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in Table 26.9.

<table>
<thead>
<tr>
<th>ODS Table Name</th>
<th>Description</th>
<th>Options</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, FIXONETIME</td>
</tr>
<tr>
<td>ParameterEstimates</td>
<td>Parameter estimates</td>
<td>Default</td>
</tr>
<tr>
<td>CovB</td>
<td>Covariances 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, DASILVA</td>
</tr>
<tr>
<td>RandomEffectsTest</td>
<td>Hausman test for random effects</td>
<td>RANONE, RANTWO</td>
</tr>
<tr>
<td>HausmanTest</td>
<td>Hausman specification test</td>
<td>HTAYLOR, AMACURDY</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 Rao score test</td>
<td>BSY</td>
</tr>
<tr>
<td>ODS Table Name</td>
<td>Description</td>
<td>Options</td>
</tr>
<tr>
<td>---------------------</td>
<td>---------------------------------------------------------------</td>
<td>------------------</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>
<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>WOOLDTTest</td>
<td>Wooldridge (2002) test for unobserved effects</td>
<td>WOOLDRIDGE02</td>
</tr>
<tr>
<td>CDTTestResults</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>DYNDIFF, DYNSYS</td>
</tr>
<tr>
<td>ARTest</td>
<td>Autoregression test for the residuals</td>
<td>DYNDIFF, DYNSYS</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 COMPARE Statement**

<table>
<thead>
<tr>
<th>StatComparisonTable</th>
<th>Comparison of model fit statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>ParameterComparisonTable</td>
<td>Comparison of model parameter</td>
</tr>
<tr>
<td></td>
<td>estimates, standard errors, and t tests</td>
</tr>
</tbody>
</table>

**ODS Tables Created by the TEST Statement**

| TestResults        | Test results                                              |
Example 26.1: The Airline Cost Data: Fixed Effects

The Christenson Associates airline data are a frequently cited data set (Greene 2000). The data measure the costs, prices of inputs, and utilization rates for six airlines from 1970 to 1984. This example analyzes the log transformations of cost (variable $C$), quantity (variable $Q$), and price (variable $PF$) and the untransformed load factor (variable $LF$). You speculate the following model,

$$
\log(C_{it}) = \alpha + \beta_1 \log(Q_{it}) + \beta_2 \log(PF_{it}) + \beta_3 LF_{it} + v_i + \epsilon_{it}
$$

where the $v_i$ are airline effects. The actual model in the original, untransformed variables is highly nonlinear:

$$
C_{it} = \exp(\alpha + \beta_3 LF_{it} + v_i + \epsilon_{it}) Q_{it}^{\beta_1} PF_{it}^{\beta_2}
$$

The following statements create the data set and perform the necessary log transformations:

```plaintext
data Airline;
  input Obs AirlineID T C Q PF LF;
  Year = T + 1969;
  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";
  label LF = "Load Factor (utilization index)";
  datalines;
  1 1 1 1140640 0.95276 106650 0.53449
  2 1 2 1215690 0.98676 110307 0.53233
  3 1 3 1309570 1.09198 110574 0.54774
  4 1 4 1511530 1.17578 121974 0.54085
  5 1 5 1676730 1.16017 196606 0.59117
  ... more lines ...
```

The following statements fit a one-way fixed-effects model:

```plaintext
proc sort data = Airline;
  by AirlineID Year;
run;

proc panel data = Airline;
  id AirlineID Year;
  model lC = lQ lPF LF / fixone printfixed;
run;
```

Output 26.1.1 provides a model and data description. There are six cross sections and 15 time points.
Example 26.1: The Airline Cost Data: Fixed Effects

Output 26.1.1 Airline Cost Data, Model Description

The PANEL Procedure
Fixed One-Way Estimates

Dependent Variable: IC (Log Transformation of Costs)

<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>
</tbody>
</table>

The R-square and degrees of freedom are shown in Output 26.1.2. The R-square statistic is nearly 1, indicating a reasonable fit. The error degrees of freedom is derived from 90 observations minus 5 cross sections, minus 4 regressors.

Output 26.1.2 Airline Cost Data, Fit Statistics

<table>
<thead>
<tr>
<th>Fit Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
</tr>
<tr>
<td>MSE</td>
</tr>
<tr>
<td>R-Square</td>
</tr>
</tbody>
</table>

The F test for fixed effects is shown in Output 26.1.3. You easily reject the null hypothesis of poolability. There are significant effects due to airlines, and it would be unreasonable to perform pooled OLS regression that ignores these effects.

Output 26.1.3 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 PRINTFIXED option in the MODEL statement provides estimates of the airline effects (which are not displayed by default). The intercept is parameterized as the fixed effect for Airline 6. The other fixed effects are differences from that base category. Quantity and fuel price have positive effects on cost, but load factors negatively affect costs. Because cost, quantity, and fuel price are log-transformed, the coefficients for quantity and price are interpreted as elasticities of cost. The coefficient for (log) fuel price is 0.417, meaning that you would associate a 10% increase in fuel price with a 4.17% increase in costs.
You suspect that there might be other factors at play, so you augment your model to include time effects. The following statements fit a two-way model, a model with both airline and time effects:

```
proc panel data = Airline;
  id AirlineID Year;
  model lC = lQ lPF LF / fixtwo printfixed;
run;
```

The $F$ test and parameter estimates for the two-way model are provided in Output 26.1.5.

**Output 26.1.4** Airline Cost Data, Parameter Estimates

| Variable | DF | Estimate | Standard Error | $t$ Value | Pr > $|t|$ | Label                      |
|----------|----|----------|----------------|-----------|-----------|---------------------------|
| CS1      | 1  | -0.08708 | 0.0842         | -1.03     | 0.3041    | Cross Sectional Effect    |
| CS2      | 1  | -0.12832 | 0.0757         | -1.69     | 0.0940    | Cross Sectional Effect    |
| CS3      | 1  | -0.29599 | 0.0500         | -5.92     | <0.0001   | Cross Sectional Effect    |
| CS4      | 1  | 0.097487 | 0.0330         | 2.95      | 0.0041    | Cross Sectional Effect    |
| CS5      | 1  | -0.06301 | 0.0239         | -2.64     | 0.0100    | Cross Sectional Effect    |
| Intercept| 1  | 9.79304  | 0.2636         | 37.15     | <0.0001   | Intercept                 |
| IQ       | 1  | 0.919293 | 0.0299         | 30.76     | <0.0001   | Log Transformation of Quantity |
| IPF      | 1  | 0.417492 | 0.0152         | 27.47     | <0.0001   | Log Transformation of Price of Fuel |
| LF       | 1  | -1.07044 | 0.2017         | -5.31     | <0.0001   | Load Factor (utilization index) |

**Output 26.1.5** Airline Cost Data, Two-Way Fixed Effects

The PANEL Procedure
Fixed Two-Way Estimates

<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>
There is an overall time trend of increasing costs. The time period of the data spans the OPEC oil embargoes and the dissolution of the Civil Aeronautics Board (CAB). These are two possible explanations for the rising costs.

A surprising result is that the fuel cost is not significant in the two-way model. If the time effects are proxies for the effect of the oil embargoes, then the effect of fuel price might be subsumed by the time effects. If the time dummy variables are proxies for the dissolution of the CAB, then the effect of load factors is not precisely estimated.

**ODS Graphics Plots**

PROC PANEL can generate ODS plots to graphically analyze the results and perform diagnostics. The following statements show how to use the PLOTS=ALL option to generate all available plots. For a complete list of options, see the section “Creating ODS Graphics” on page 1893.

```plaintext
ods graphics on;
proc panel data = Airline;
   id AirlineID Year;
   model lC = lQ lPF LF / fixtwo plots = all;
run;
```

Specifying PLOTS=ALL produces two panels of plots, shown in Output 26.1.6 and Output 26.1.7.
Output 26.1.6 Airline Cost Data, Diagnostic Panel 1

Fit Diagnostics for IC

Observations 90  MSE 0.002639  Model DF 67
The following statements demonstrate how to use the UNPACK option to unpack the panels into single plots, and how to use the ONLY option to select only a surface plot of residuals:

```plaintext
proc panel data = Airline;
   id AirlineID Year;
   model lC = lQ lPF LF / fixtwo plots(unpack only) = residsurface;
run;
```

The unpacked residual-surface plot is shown in Output 26.1.8.
Example 26.2: Analyzing Demand for Liquid Assets: Random Effects

Feige (1964) provides data on the demand for liquid assets. The data are for six states and the District of Columbia (CA, DC, FL, IL, NY, TX, and WA) and were collected each year from 1949 to 1959. All variables are log-transformed.

The following statements create the Assets data set:

```plaintext
data Assets;
  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
```

The output shows the surface plot of the residuals for IC, indicating the trend over the years for each state.
Example 26.2: Analyzing Demand for Liquid Assets: Random Effects

The data contain per capita consumptions for three liquid assets: demand deposits such as checking, time deposits, and savings and loan (S&L) shares. You posit a linear model for per capita demand deposits, with random effects for states.

The following statements fit a one-way random-effects model:

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

proc panel data = Assets;
   id state year;
   model d = y rd rt rs / ranone;
run;
```

The regression results are provided in Output 26.2.1.

The “Variance Component Estimates” table provides the estimated variance of the cross-sectional (state) effects and the variance of the observation-level errors. A majority of the overall error variance can be attributed to differences between states, not differences within states.

The “Hausman Test for Random Effects” table shows the result of a Hausman specification test. The null hypothesis is that state effects can be treated as random (random-effects model) and that they do not need to be estimated directly (fixed-effects model). The test results favor the random-effects specification that is used to generate this output.

### Output 26.2.1 Demand for Demand Deposits, One-Way Random-Effects Model

#### The PANEL Procedure
**Fuller and Battese Variance Components (RanOne)**

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

<table>
<thead>
<tr>
<th>Fit Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SSE</strong></td>
</tr>
<tr>
<td><strong>DFE</strong></td>
</tr>
<tr>
<td><strong>MSE</strong></td>
</tr>
<tr>
<td><strong>Root MSE</strong></td>
</tr>
<tr>
<td><strong>R-Square</strong></td>
</tr>
</tbody>
</table>

##### Variance Component Estimates

| Variance Component for Cross Sections | 0.029067 |
| Variance Component for Error | 0.00134 |
The parameter estimate for the variable $Y$ is greater than 1, indicating that demand is elastic to income—income has a more than proportional positive association with the demand for demand deposits. The coefficient on the variable $RD$ indicates that demand deposits increase significantly as the service charge is reduced.

The variables $RT$ and $RS$ represent positive aspects of competing products, and you would expect these variables to affect demand negatively. The coefficient for $RS$ meets that expectation, but the coefficient for $RT$ is not significant.

The previous analysis used the default Fuller-Battese method to estimate the variance components. The PANEL procedure supports three other methods, and you might be interested in how use of the different methods affects the analysis.

The following statements fit the model by using all four methods and include a COMPARE statement to compare the results:

```sas
proc panel data = Assets;
  id state year;
  wh: model d = y rd rt rs / ranone vcomp = wh;
  wk: model d = y rd rt rs / ranone vcomp = wk;
  fb: model d = y rd rt rs / ranone vcomp = fb;
  nl: model d = y rd rt rs / ranone vcomp = nl;
  compare / mstat(varcs varerr);
run;
```

The tables that the COMPARE statement produces are shown in Output 26.2.2.
Output 26.2.2 One-Way versus Two-Way Random Effects, Assets Data

The PANEL Procedure
Model Comparison

Dependent Variable: d (Per Capita Demand Deposits)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>WH RanOne</th>
<th>WK RanOne</th>
<th>FB RanOne</th>
<th>NL RanOne</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var due to Cross Sections</td>
<td>0.0315</td>
<td>0.0315</td>
<td>0.0291</td>
<td>0.0327</td>
</tr>
<tr>
<td>Var due to Error</td>
<td>0.000107</td>
<td>0.001340</td>
<td>0.001340</td>
<td>0.001149</td>
</tr>
</tbody>
</table>

Comparison of Model Parameter Estimates

<table>
<thead>
<tr>
<th>Variable</th>
<th>WH RanOne</th>
<th>WK RanOne</th>
<th>FB RanOne</th>
<th>NL RanOne</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1.472425</td>
<td>-1.723092</td>
<td>-1.742581</td>
<td>-1.680406</td>
</tr>
<tr>
<td>Std Err</td>
<td>0.719067</td>
<td>0.681184</td>
<td>0.680541</td>
<td>0.682676</td>
</tr>
<tr>
<td>y</td>
<td>1.117252</td>
<td>1.145844</td>
<td>1.148051</td>
<td>1.141001</td>
</tr>
<tr>
<td>Std Err</td>
<td>0.099799</td>
<td>0.099776</td>
<td>0.099761</td>
<td>0.099802</td>
</tr>
<tr>
<td>rd</td>
<td>-0.245861</td>
<td>-0.272995</td>
<td>-0.275135</td>
<td>-0.268325</td>
</tr>
<tr>
<td>Std Err</td>
<td>0.052260</td>
<td>0.051445</td>
<td>0.051372</td>
<td>0.051600</td>
</tr>
<tr>
<td>rt</td>
<td>0.029227</td>
<td>0.033397</td>
<td>0.033718</td>
<td>0.032692</td>
</tr>
<tr>
<td>Std Err</td>
<td>0.028570</td>
<td>0.029416</td>
<td>0.029485</td>
<td>0.029266</td>
</tr>
<tr>
<td>rs</td>
<td>-0.414540</td>
<td>-0.410731</td>
<td>-0.410361</td>
<td>-0.411500</td>
</tr>
<tr>
<td>Std Err</td>
<td>0.117486</td>
<td>0.119968</td>
<td>0.120160</td>
<td>0.119548</td>
</tr>
</tbody>
</table>

You conclude that how you estimate variance components has little bearing on the regression results.

It is possible that there are time random effects in addition to random effects for states. To explore this possibility, you fit a two-way random-effects model and again use a COMPARE statement to conveniently compare the one- and two-way models:

```plaintext
proc panel data = Assets;
   id state year;
   model d = y rd rt rs / ranone rantwo;
   compare;
run;
```

The model comparison table is shown in Output 26.2.3. Although the parameter estimates differ somewhat, your interpretation of the effects on demand remains unchanged.

Cornwell and Rupert (1988) analyze data from the Panel Study of Income Dynamics (PSID), an income study of 595 individuals over the seven-year period 1976–1982. Of particular interest is the effect of additional schooling on wages. The analysis here replicates that of Baltagi (2013, sec. 7.5), where it is concluded that covariate correlation with individual effects makes a standard random-effects model inadequate.

The following statements create the PSID data set:

data psid;
  input id t lwage wks south smsa ms exp exp2 occ ind union fem blk ed;
  label id = 'Person ID'
    t = 'Time'
    lwage = 'Log(wages)'
    wks = 'Weeks worked'
    south = '1 if resides in the South'
    smsa = '1 if resides in SMSA'
    ms = '1 if married'
    exp = 'Years full-time experience'
    exp2 = 'exp squared'
    occ = '1 if blue-collar occupation'
    ind = '1 if manufacturing'
    union = '1 if union contract'
    fem = '1 if female'
    blk = '1 if black'
    ed = 'Years of education';
datalines;
  1 1 5.5606799126 32 1 0 1 3 9 0 0 0 0 0 9
  1 2 5.7203102112 43 1 0 1 4 16 0 0 0 0 0 9
  1 3 5.9964499474 40 1 0 1 5 25 0 0 0 0 0 9

1  4  5.9964499474 39  1  0  1  6  36  0  0  0  0  0  9
1  5  6.0614600182 42  1  0  1  7  49  0  1  0  0  0  9
1  6  6.1737899780 35  1  0  1  8  64  0  1  0  0  0  9
1  7  6.2441701889 32  1  0  1  9  81  0  1  0  0  0  9
2  1  6.1633100510 34  0  0  1  30  900 1  0  0  0  0  11
2  2  6.2146100998 27  0  0  1  31  961 1  0  0  0  0  11
2  3  6.2634000778 33  0  0  1  32 1024 1  1  1  0  0  11
2  4  6.5439100266 30  0  0  1  33 1089 1  1  0  0  0  11
2  5  6.6970300674 30  0  0  1  34 1156 1  1  0  0  0  11
2  6  6.7912201881 37  0  0  1  35 1225 1  1  0  0  0  11
2  7  6.8156399727 30  0  0  1  36 1296 1  1  0  0  0  11

... more lines ...

You begin by fitting a one-way random-effects model:

```bash
proc sort data=psid;
  by id t;
run;

proc panel data=psid;
  id id t;
  model lwage = wks south smsa ms exp exp2 occ
        ind union fem blk ed / ranone;
run;
```

The output is shown in Output 26.3.1. The coefficient on the variable ED (which represents years of education) estimates that an additional year of schooling is associated with about a 10.7% increase in wages. However, the results of the Hausman test for random effects show a serious violation of the random-effects assumptions, namely that the regressors are independent of both error components.

**Output 26.3.1 One-Way Random-Effects Estimation**

**The PANEL Procedure**

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

**Dependent Variable: lwage (Log(wages))**

<table>
<thead>
<tr>
<th>Model Description</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
<td>RanOne</td>
</tr>
<tr>
<td>Number of Cross Sections</td>
<td>595</td>
</tr>
<tr>
<td>Time Series Length</td>
<td>7</td>
</tr>
</tbody>
</table>

**Variance Component Estimates**

| Variance Component for Cross Sections | 0.100553 |
| Variance Component for Error          | 0.023102 |

**Hausman Test for Random Effects**

| Coefficients DF m Value Pr > m |
|-------------------------------|-----------------------------|
| 9 9 5288.98 <.0001            |
An alternative could be a fixed-effects (FIXONE option) model, but that would not permit estimation of the coefficient for \( ED \), which does not vary within individuals. A compromise is the Hausman-Taylor model, for which you stipulate a set of covariates that are correlated with the individual effects (but uncorrelated with the observation-level errors). You specify the correlated variables in the CORRELATED= option in the INSTRUMENTS statement:

```plaintext
proc panel data=psid;
  id id t;
  instruments correlated = (wks ms exp exp2 union ed);
  model lwage = wks south_smsa ms exp exp2 occ
    ind union fem blk ed / htaylor;
run;
```

The results are shown in Output 26.3.2. The table of parameter estimates has an added column, Type, that identifies which regressors are assumed to be correlated with individual effects (C) and which regressors do not vary within cross sections (TI). It was stated previously that the Hausman-Taylor model is a compromise between fixed-effects and random-effects models, and you can think of the compromise this way: You want to fit a random-effects model, but the correlated (C) variables make that model invalid. So you revert to the consistent fixed-effects model, but then the time-invariant (TI) variables are the problem because they will be dropped from that model. The solution is to use the Hausman-Taylor estimator.

The estimation results show that an additional year of schooling is now associated with a 13.8% increase in wages. Also presented is a Hausman test that compares this model to the fixed-effects model. As was the case previously when you fit the random-effects model, you can think of the Hausman test as a referendum on the assumptions that you are making. For this estimation, it seems that your choice of variables to treat as correlated is adequate. It also seems to be true that any correlation is with the individual-level effects, not the observation-level errors.
Output 26.3.2 Hausman-Taylor Estimation

The PANEL Procedure
Hausman and Taylor Model for Correlated Individual Effects (HTaylor)

Dependent Variable: lwage (Log(wages))

<table>
<thead>
<tr>
<th>Variance Component Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variance Component for Cross Sections</td>
</tr>
<tr>
<td>Variance Component for Error</td>
</tr>
</tbody>
</table>

Hausman Test against Fixed Effects

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>DF</th>
<th>m Value</th>
<th>Pr &gt; m</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9</td>
<td>5.26</td>
<td>0.1539</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Intercept</td>
</tr>
<tr>
<td>wks</td>
</tr>
<tr>
<td>south</td>
</tr>
<tr>
<td>smsa</td>
</tr>
<tr>
<td>ms</td>
</tr>
<tr>
<td>exp</td>
</tr>
<tr>
<td>exp2</td>
</tr>
<tr>
<td>occ</td>
</tr>
<tr>
<td>ind</td>
</tr>
<tr>
<td>union</td>
</tr>
<tr>
<td>fem</td>
</tr>
<tr>
<td>blk</td>
</tr>
<tr>
<td>ed</td>
</tr>
</tbody>
</table>

C: correlated with the individual effects
TI: constant (time-invariant) within cross sections

At its core, the Hausman-Taylor estimator is an instrumental variables regression, where the instruments are derived from regressors that are assumed to be uncorrelated with the individual effects. Technically, it is the cross-sectional means of these variables that need to be uncorrelated, not the variables themselves.

The Amemiya-MaCurdy model is a close relative of the Hausman-Taylor model. The only difference between the two is that the Amemiya-MaCurdy model makes the added assumption that the regressors (and not just their means) are uncorrelated with the individual effects. By making that assumption, the Amemiya-MaCurdy model can take advantage of a more efficient set of instrumental variables.

The following statements fit the Amemiya-MaCurdy model:

```
proc panel data=psid;
   id id t;
   instruments correlated = (wks ms exp exp2 union ed);
   model lwage = wks south smsa ms exp exp2 occ
                  ind union fem blk ed / amacurdy;
run;
```
The results are shown in Output 26.3.3. Little is changed from the Hausman-Taylor model. The Hausman test compares the Amemiya-MaCurdy model to the Hausman-Taylor model and shows that the one additional assumption is acceptable. You even gain a bit of efficiency in the process: compare the standard deviations of the coefficient on the variable ED from both models.

**Output 26.3.3 Amemiya-MaCurdy Estimation**

The PANEL Procedure
Amemiya and MaCurdy Model for Correlated Individual Effects (AMaCurdy)

<table>
<thead>
<tr>
<th>Dependent Variable: Iwage (Log(wages))</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Variance Component Estimates</strong></td>
</tr>
<tr>
<td>Variance Component for Cross Sections 0.886993</td>
</tr>
<tr>
<td>Variance Component for Error 0.023044</td>
</tr>
</tbody>
</table>

<p>| <strong>Hausman Test against Hausman-Taylor</strong> |</p>
<table>
<thead>
<tr>
<th>Coefficients</th>
<th>DF</th>
<th>m Value</th>
<th>Pr &gt; m</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>13</td>
<td>14.67</td>
<td>0.3287</td>
</tr>
</tbody>
</table>

**Parameter Estimates**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Type</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>Intercept</td>
<td></td>
<td>1</td>
<td>2.927338</td>
<td>0.2751</td>
<td>10.64</td>
<td>&lt;.0001</td>
<td>Intercept</td>
</tr>
<tr>
<td>wks</td>
<td>C</td>
<td>1</td>
<td>0.000838</td>
<td>0.000599</td>
<td>1.40</td>
<td>0.1622</td>
<td>Weeks worked</td>
</tr>
<tr>
<td>south</td>
<td></td>
<td>1</td>
<td>0.007282</td>
<td>0.0319</td>
<td>0.23</td>
<td>0.8197</td>
<td>1 if resides in the South</td>
</tr>
<tr>
<td>smsa</td>
<td></td>
<td>1</td>
<td>-0.04195</td>
<td>0.0189</td>
<td>-2.21</td>
<td>0.0269</td>
<td>1 if resides in SMSA</td>
</tr>
<tr>
<td>ms</td>
<td>C</td>
<td>1</td>
<td>-0.03009</td>
<td>0.0190</td>
<td>-1.59</td>
<td>0.1127</td>
<td>1 if married</td>
</tr>
<tr>
<td>exp</td>
<td>C</td>
<td>1</td>
<td>0.11297</td>
<td>0.00247</td>
<td>45.76</td>
<td>&lt;.0001</td>
<td>Years full-time experience</td>
</tr>
<tr>
<td>exp2</td>
<td>C</td>
<td>1</td>
<td>-0.00042</td>
<td>0.000055</td>
<td>-7.72</td>
<td>&lt;.0001</td>
<td>exp squared</td>
</tr>
<tr>
<td>occ</td>
<td></td>
<td>1</td>
<td>-0.02085</td>
<td>0.0138</td>
<td>-1.51</td>
<td>0.1299</td>
<td>1 if blue-collar occupation</td>
</tr>
<tr>
<td>ind</td>
<td></td>
<td>1</td>
<td>0.013629</td>
<td>0.0152</td>
<td>0.89</td>
<td>0.3709</td>
<td>1 if manufacturing</td>
</tr>
<tr>
<td>union</td>
<td>C</td>
<td>1</td>
<td>0.032475</td>
<td>0.0149</td>
<td>2.18</td>
<td>0.0293</td>
<td>1 if union contract</td>
</tr>
<tr>
<td>fem</td>
<td>TI</td>
<td>1</td>
<td>-0.13201</td>
<td>0.1266</td>
<td>-1.04</td>
<td>0.2972</td>
<td>1 if female</td>
</tr>
<tr>
<td>blk</td>
<td>TI</td>
<td>1</td>
<td>-0.2859</td>
<td>0.1555</td>
<td>-1.84</td>
<td>0.0660</td>
<td>1 if black</td>
</tr>
<tr>
<td>ed</td>
<td>C</td>
<td>TI</td>
<td>0.137205</td>
<td>0.0206</td>
<td>6.67</td>
<td>&lt;.0001</td>
<td>Years of education</td>
</tr>
</tbody>
</table>

**C: correlated with the individual effects**
**TI: constant (time-invariant) within cross sections**

Finally, you should realize that the Hausman-Taylor and Amemiya-MaCurdy estimators are not cure-alls for correlated individual effects. Estimation tacitly relies on the uncorrelated regressors being sufficient to predict the correlated regressors. Otherwise, you run into the problem of weak instruments. If you have weak instruments, you will obtain biased estimates that have very large standard errors. However, that does not seem to be the case here.

**Example 26.4: Cigarette Sales Data: Dynamic Panel Estimation**

Consider a dynamic panel demand model for cigarette sales that illustrates the methods described in the section “Dynamic Panel Estimation (DYNDIFF and DYNYSYS Options)” on page 1855. The data are from
Example 26.4: Cigarette Sales Data: Dynamic Panel Estimation

a panel of 46 American states over the period 1963–1992. The dependent variable is the logarithm of per capita cigarette sales (variable LSales). Other factors that were measured include the log of price (LPrice), the log of disposable income (LDisp), and the log of minimum price in adjoining states (LMin). For a full description of the data, see Baltagi (2013, sec. 8.9).

The following statements create the Cigar data set:

```sas
data Cigar;
  input State Year Price Pop Pop_16 Cpi Disp Sales Min;
  LSales = log(Sales);
  LPrice = log(Price);
  LDisp = log(Disp);
  LMin = log(Min);
  label State = 'State abbreviation'
           Year = 'Year'
           LSales = 'Log cigarette sales in packs per capita'
           LPrice = 'Log price per pack of cigarettes'
           LDisp = 'Log per capita disposable income'
           LMin = 'Log minimum price in adjoining states per pack of cigarettes';
datalines;
  ... more lines ...
```

You posit a panel model for cigarette sales that contains fixed effects for states. Because you believe that the data are insufficient to explain all possible shocks in yearly sales, you include lagged sales in the model as a regressor. By construction, lagged sales are an endogenous regressor, and you thus specify dynamic panel estimation by using the DYNDIFF option. The following statements fit the model:

```sas
proc sort data=Cigar;
  by State Year;
run;

proc panel data=Cigar;
  id State Year;
  model LSales = LPrice LDisp LMin / dyndiff;
run;
```

The results are shown in Output 26.4.1. Note that it was not necessary to explicitly include lagged sales on the right-hand side of the model; PROC PANEL generated it for you. The coefficient on lagged sales is 0.732, indicating a high degree of autocorrelation in the dependent variable. When cigarette sales are unusually high or low because of unforeseen circumstances, the effects tend to linger for several years. The results also show that demand is highly elastic to price.
Output 26.4.1  Dynamic Panel Estimation for Cigarette Sales

The PANEL Procedure
Dynamic Panel Estimation by First-Differences  GMM

Dependent Variable: LSales (Log cigarette sales in packs per capita)

<table>
<thead>
<tr>
<th>Model Description</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
<td>DynDiff</td>
</tr>
<tr>
<td>Number of Cross Sections</td>
<td>46</td>
</tr>
<tr>
<td>Time Series Length</td>
<td>30</td>
</tr>
<tr>
<td>GMM Stage</td>
<td>1</td>
</tr>
<tr>
<td>GMM Bandwidth</td>
<td>30</td>
</tr>
<tr>
<td>Number of Instruments</td>
<td>410</td>
</tr>
<tr>
<td>Variance Estimation</td>
<td>GMM</td>
</tr>
</tbody>
</table>

Fit Statistics

\[
\begin{align*}
\text{SSE} & = 3.1373 \\
\text{DFE} & = 1283 \\
\text{MSE} & = 0.0024 \\
\text{Root MSE} & = 0.0494
\end{align*}
\]

Sargan Test

\[
\begin{align*}
\text{DF Statistic} & = 712.45 \\
\text{Prob > ChiSq} & = <.0001
\end{align*}
\]

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>DF</td>
<td>Estimate</td>
<td>Error</td>
<td>t Value</td>
</tr>
<tr>
<td>Intercept</td>
<td>1</td>
<td>0.769092</td>
<td>0.0658</td>
<td>11.69</td>
</tr>
<tr>
<td>LSales (Lag 1)</td>
<td>1</td>
<td>0.732212</td>
<td>0.0178</td>
<td>41.07</td>
</tr>
<tr>
<td>LPrice</td>
<td>1</td>
<td>-0.26328</td>
<td>0.0255</td>
<td>-10.31</td>
</tr>
<tr>
<td>LDisp</td>
<td>1</td>
<td>0.166116</td>
<td>0.0105</td>
<td>15.88</td>
</tr>
<tr>
<td>LMin</td>
<td>1</td>
<td>0.032726</td>
<td>0.0233</td>
<td>1.40</td>
</tr>
</tbody>
</table>

AR(m) Test

\[
\begin{align*}
\text{Lag Statistic} & = 2.47 \\
\text{Pr > |Statistic|} & = 0.0134
\end{align*}
\]

Included in Output 26.4.1 are two diagnostic measures. The first, a Sargan test, is a test of the validity of the moment conditions that are conferred by the GMM instruments that were used. The \( p \)-value indicates that the moment conditions are not valid and that you should probably look for a set of instruments other than the default set provided by PROC PANEL.

The second diagnostic test is the AR(m) test for autocorrelation in the residuals. In well-fitting dynamic panel models, you expect to see some autocorrelation of lag 1, but any autocorrelation at higher lags indicates a poor fit. The autocorrelation at lag 2 is significant, leading you to seek a better-fitting alternative.

One possible explanation for the poor fit is that, by default, PROC PANEL uses the one-step generalized method of moments (GMM). One-step GMM is known for being too reliant on the assumption that the residuals from the difference equations are not serially correlated. An alternative is two-step GMM, which instead uses a data-driven variance matrix for the differenced residuals.
The following statements fit the model by two-step GMM:

```plaintext
proc panel data=Cigars;
    id State Year;
    instruments constant depvar diffeq=(LPrice LDisp LMin);
    model lSales = LPrice LDisp LMin / dyndiff twostep biascorrected;
run;
```

The code includes an INSTRUMENTS statement that, for demonstration purposes, reproduces the default instrument set. That set includes the following:

- a constant (keyword CONSTANT)
- GMM-style instruments based on the dependent variable, LSales (keyword DEPVAR)
- standard instruments for the exogenous regressors LPrice, LDisp, and LMin (DIFFEQ= option)

The code also includes the BIASCORRECTED option, which produces bias-corrected standard errors according to the method of Windmeijer (2005).

The results are shown in Output 26.4.2. The coefficients do not change much, but the standard errors are now more reliable. The model diagnostic tests indicate a better fit, although you should use caution when interpreting Sargan test results. Sargan tests lack power when the number of instruments is large, and their distributional properties come into question under conditions that favor either robust or bias-corrected standard errors.

**Output 26.4.2** Dynamic Panel Estimation by Two-Step GMM

**The PANEL Procedure**
**Dynamic Panel Estimation by First-Differences**
**GMM**

**Dependent Variable:** LSales (Log cigarette sales in packs per capita)

<table>
<thead>
<tr>
<th>Model Description</th>
<th>DynDiff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimation Method</td>
<td></td>
</tr>
<tr>
<td>Number of Cross Sections</td>
<td>46</td>
</tr>
<tr>
<td>Time Series Length</td>
<td>30</td>
</tr>
<tr>
<td>GMM Stage</td>
<td>2</td>
</tr>
<tr>
<td>GMM Bandwidth</td>
<td>30</td>
</tr>
<tr>
<td>Number of Instruments</td>
<td>410</td>
</tr>
<tr>
<td>Variance Estimation</td>
<td>Bias-corrected</td>
</tr>
</tbody>
</table>

**Fit Statistics**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
<td>3.1348</td>
</tr>
<tr>
<td>DFE</td>
<td>1283</td>
</tr>
<tr>
<td>MSE</td>
<td>0.0024</td>
</tr>
<tr>
<td>Root MSE</td>
<td>0.0494</td>
</tr>
</tbody>
</table>

**Sargan Test**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>DF</td>
<td>41</td>
</tr>
<tr>
<td>Prob &gt; ChiSq</td>
<td>0.2920</td>
</tr>
</tbody>
</table>
The previous estimation treats regressors such as \( \text{LPrice} \) as exogenous. If you believe that price is endogenous, you can create GMM-style instruments for \( \text{LPrice} \) to replace the default standard instruments.

The following statements fit the model by using GMM-style instruments for \( \text{LPrice} \):

```
proc panel data=Cigar;
  id State Year;
  instruments constant depvar diffeq=(LDisp LMin) diffend=(LPrice);
  model lSales = LPrice LDisp LMin / dyndiff twostep biascorrected;
run;
```

The results are shown in Output 26.4.3. Treating \( \text{LPrice} \) as endogenous greatly increases the number of instruments. Although this is not the case here, when the number of instruments is so large that it makes estimation infeasible, you can limit the number of instruments by specifying the MAXBAND= option in the INSTRUMENTS statement.

```
Output 26.4.3  Dynamic Panel Estimation, Custom Instrument Set
```

**The PANEL Procedure**

**Dynamic Panel Estimation by First-Differences  GMM**

**Dependent Variable: LSales (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>GMM Stage</td>
</tr>
<tr>
<td>GMM Bandwidth</td>
</tr>
<tr>
<td>Number of Instruments</td>
</tr>
<tr>
<td>Variance Estimation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fit Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSE</td>
</tr>
<tr>
<td>DFE</td>
</tr>
<tr>
<td>MSE</td>
</tr>
<tr>
<td>Root MSE</td>
</tr>
</tbody>
</table>
```
Example 26.5: Using the FLATDATA Statement

Sometimes data sets are stored in compressed (or wide) form, where each record contains all observations for the entire cross section. Although the PANEL procedure requires data in uncompressed (long) form, sometimes it is easier to create new variables or summary statistics if the data are in wide form.

To illustrate, suppose you have a simulated data set that contains 20 cross sections measured over six time periods. Each time period has values for dependent and independent variables, \( Y_1, \ldots, Y_6 \) and \( X_1, \ldots, X_6 \). The \( cs \) and num variables are constant across each cross section.

The observations for the first five cross sections along with other variables are shown in Output 26.5.1. In this example, \( i \) represents the cross section. The time period is identified by the subscript of the \( Y \) and \( X \) variables, which ranges from 1 to 6.

Output 26.5.1  Compressed Data Set

<table>
<thead>
<tr>
<th>Obs</th>
<th>i</th>
<th>cs</th>
<th>num</th>
<th>X_1</th>
<th>X_2</th>
<th>X_3</th>
<th>X_4</th>
<th>X_5</th>
<th>X_6</th>
<th>Y_1</th>
<th>Y_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>CS1</td>
<td>-1.56058</td>
<td>0.40268</td>
<td>0.91951</td>
<td>0.69482</td>
<td>-2.28899</td>
<td>-1.32762</td>
<td>1.92348</td>
<td>2.30418</td>
<td>2.11850</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>CS2</td>
<td>0.30989</td>
<td>1.01950</td>
<td>-0.04699</td>
<td>-0.96695</td>
<td>-1.08345</td>
<td>-0.05180</td>
<td>0.30266</td>
<td>4.50982</td>
<td>3.73887</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>CS3</td>
<td>0.85054</td>
<td>0.60325</td>
<td>0.71154</td>
<td>0.66168</td>
<td>-0.66823</td>
<td>-1.87550</td>
<td>0.55065</td>
<td>4.07276</td>
<td>4.89621</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>CS4</td>
<td>-0.18885</td>
<td>-0.64946</td>
<td>-1.23355</td>
<td>0.04554</td>
<td>-0.24996</td>
<td>0.09685</td>
<td>-0.92771</td>
<td>2.40304</td>
<td>1.48182</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>CS5</td>
<td>-0.04761</td>
<td>-0.79692</td>
<td>0.63445</td>
<td>-2.23539</td>
<td>-0.37629</td>
<td>-0.82212</td>
<td>-0.70566</td>
<td>3.58092</td>
<td>6.08917</td>
</tr>
</tbody>
</table>

Output 26.4.3  continued

<table>
<thead>
<tr>
<th>Sargan Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>DF</td>
</tr>
<tr>
<td>41</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Intercep</td>
</tr>
<tr>
<td>LSales (Lag 1)</td>
</tr>
<tr>
<td>LPrice</td>
</tr>
<tr>
<td>LDisp</td>
</tr>
<tr>
<td>LMin</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AR(m) Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lag</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
</tbody>
</table>
When the data are in this form, it is easy to create other variables that are combinations of the existing variables. For example, you can calculate the within-cross-section mean of $X$ by simply summing across the $X_i$ variables and dividing by six. It is easier to perform this kind of data manipulation when the data are in compressed (wide) form instead of uncompressed (long) form.

On the other hand, the PANEL procedure cannot work directly with the data in wide form. You can use the FLATDATA statement to transform wide data into long form “on the fly” for performing a panel data analysis. You can also use the OUT= option to output the transformed data to a new data set, to use for further analysis.

The following code reshapes the data and performs fixed-effects estimation:

```sparql
proc panel data=flattest;
  flatdata indid=i tsname="t" base=(X Y)
    keep=( cs num seed ) / out=flat_out;
  id i t;
  model y = x / fixone noint;
run;
```

The first six observations in the uncompressed (long) data set and the results for the one-way fixed-effects model are shown in Output 26.5.2 and Output 26.5.3, respectively.

### Output 26.5.2 Uncompressed Data Set

<table>
<thead>
<tr>
<th>Obs</th>
<th>t</th>
<th>X</th>
<th>Y</th>
<th>CS</th>
<th>NUM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0.40268</td>
<td>2.30418</td>
<td>CS1</td>
<td>-1.56058</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0.91951</td>
<td>2.11850</td>
<td>CS1</td>
<td>-1.56058</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0.69482</td>
<td>2.66009</td>
<td>CS1</td>
<td>-1.56058</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>-2.28999</td>
<td>-4.94104</td>
<td>CS1</td>
<td>-1.56058</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>-1.32762</td>
<td>-0.83053</td>
<td>CS1</td>
<td>-1.56058</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>1.92348</td>
<td>5.01359</td>
<td>CS1</td>
<td>-1.56058</td>
</tr>
</tbody>
</table>

### Output 26.5.3 Estimation with the FLATDATA Statement

**The PANEL Procedure**

**Fixed One-Way Estimates**

**Dependent Variable: Y**

<table>
<thead>
<tr>
<th>Parameter Estimates</th>
<th>Standard Error</th>
<th>t Value</th>
<th>Pr &gt;</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>2.010753</td>
<td>0.1217</td>
<td>16.52</td>
<td>&lt;.0001</td>
</tr>
</tbody>
</table>

Now, suppose you have long data that you want to reshape into wide form. The following DATA step performs this task:

```sparql
data wide;
  set flat_out;
  by i;
  keep i num cs X_1-X_6 Y_1-Y_6;
  retain X_1-X_6 Y_1-Y_6;
  array ax(1:6) X_1-X_6;
```
array ay(1:6) Y_1-Y_6;
if first.i then do;
   do j = 1 to 6;
      ax(j) = 0;
      ay(j) = 0;
   end;
end;
ax(t) = X;
ay(t) = Y;
if last.i then output;
run;

As a check, Output 26.5.4 lists the newly compressed data, which match the original data from this example.

**Output 26.5.4** Repressed Data Set

<table>
<thead>
<tr>
<th>Obs</th>
<th>I</th>
<th>CS</th>
<th>NUM</th>
<th>X_1</th>
<th>X_2</th>
<th>X_3</th>
<th>X_4</th>
<th>X_5</th>
<th>X_6</th>
<th>Y_1</th>
<th>Y_2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>CS1</td>
<td>0.40268</td>
<td>0.91951</td>
<td>0.69482</td>
<td>-2.28899</td>
<td>-1.32762</td>
<td>1.92348</td>
<td>2.30418</td>
<td>2.11850</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>CS2</td>
<td>0.30989</td>
<td>1.01950</td>
<td>-0.04699</td>
<td>-0.96695</td>
<td>-1.08345</td>
<td>-0.05180</td>
<td>0.30266</td>
<td>4.50982</td>
<td>3.73887</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>CS3</td>
<td>0.85054</td>
<td>0.60325</td>
<td>0.71154</td>
<td>0.66168</td>
<td>-0.66823</td>
<td>-1.87550</td>
<td>0.55065</td>
<td>4.07276</td>
<td>4.89621</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>CS4</td>
<td>-0.18885</td>
<td>-0.64946</td>
<td>-1.23355</td>
<td>0.04554</td>
<td>-0.24996</td>
<td>0.09685</td>
<td>-0.92771</td>
<td>2.40304</td>
<td>1.48182</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>CS5</td>
<td>-0.04761</td>
<td>-0.79692</td>
<td>0.63445</td>
<td>-2.23539</td>
<td>-0.37629</td>
<td>-0.82212</td>
<td>-0.70566</td>
<td>3.58092</td>
<td>6.08917</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Obs</th>
<th>Y_3</th>
<th>Y_4</th>
<th>Y_5</th>
<th>Y_6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.66009</td>
<td>-4.94104</td>
<td>0.83053</td>
<td>5.01359</td>
</tr>
<tr>
<td>2</td>
<td>1.44984</td>
<td>-1.02996</td>
<td>2.78260</td>
<td>1.73856</td>
</tr>
<tr>
<td>3</td>
<td>3.99470</td>
<td>1.03437</td>
<td>0.54598</td>
<td>5.01460</td>
</tr>
<tr>
<td>4</td>
<td>2.70579</td>
<td>3.82672</td>
<td>4.01117</td>
<td>1.97639</td>
</tr>
<tr>
<td>5</td>
<td>3.08249</td>
<td>4.26605</td>
<td>3.65452</td>
<td>0.81826</td>
</tr>
</tbody>
</table>

**References**


Subject Index

Amemiya-MaCurdy estimation, 1855
   PANEL procedure, 1855
between-groups regression, 1840
   PANEL procedure, 1840
BY groups
   PANEL procedure, 1811
Da Silva method
   PANEL procedure, 1851
fixed-effects model
   fixed-effects test, 1873
   one-way first differencing, 1842
   two-way first differencing, 1843
fixed-effects test
   fixed-effects model, 1873
Fuller-Battese variance components, 1844, 1848
generalized least squares
   PANEL procedure, 1850
GMM: Arellano-Bond estimator
   panel GMM, 1855
HAC option
   PANEL procedure, 1870
Hausman-Taylor estimation, 1853
   PANEL procedure, 1853
HCCME= option
   PANEL procedure, 1866
heteroscedasticity- and autocorrelation-consistent covariance matrices, 1870
heteroscedasticity-corrected covariance matrices, 1866
ID variables
   PANEL procedure, 1815
linear hypothesis testing, 1865
   PANEL procedure, 1865
linear restrictions
   PANEL procedure, 1865
Nerlove variance components, 1845, 1848
one-way first differencing
   fixed-effects model, 1842
one-way fixed-effects by first differencing, 1842
   PANEL procedure, 1842
one-way fixed-effects model, 1840
   PANEL procedure, 1840
one-way random-effects model, 1843
   PANEL procedure, 1843
options summary
   MODEL statement (PANEL), 1818
output data sets
   PANEL procedure, 1893–1895
output table names
   PANEL procedure, 1896
panel GMM, 1855
   GMM: Arellano-Bond estimator, 1855
PANEL procedure
   Amemiya-MaCurdy estimation, 1855
   between-groups regression, 1840
   BY groups, 1811
   Da Silva method, 1851
   generalized least squares, 1850
   HAC option, 1870
   Hausman-Taylor estimation, 1853
   HCCME= option, 1866
   ID variables, 1815
   linear hypothesis testing, 1865
   linear restrictions, 1865
   one-way fixed-effects by first differencing, 1842
   one-way fixed-effects model, 1840
   one-way random-effects model, 1843
   output data sets, 1893–1895
   output table names, 1896
   Parks method, 1849
   pooled regression, 1840
   predicted values, 1835
   printed output, 1895
   R-square measure, 1872
   residuals, 1835
   tests for random effects, 1873
   two-way fixed-effects by first differencing, 1843
   two-way fixed-effects model, 1841
   two-way random-effects model, 1845
   Zellner’s two-stage method, 1851
Parks method
   PANEL procedure, 1849
   pooled regression, 1840
   predicted values
      PANEL procedure, 1835
   printed output
      PANEL procedure, 1895
   R-square measure
residuals
PANEL procedure, 1835
restricted estimation, 1865
selecting variance components, 1845
tests for random effects
PANEL procedure, 1873
two-stage least squares (2SLS)
regression, 1853, 1855
two-way first differencing
fixed-effects model, 1843
two-way fixed-effects by first differencing, 1843
PANEL procedure, 1843
two-way fixed-effects model, 1841
PANEL procedure, 1841
two-way random-effects model, 1845
PANEL procedure, 1845
Wallace-Hussain variance components, 1843, 1846
Wansbeek-Kapteyn variance components, 1844, 1847
Zellner’s two-stage method
PANEL procedure, 1851
## Syntax Index

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL option</td>
<td>TEST statement (PANEL), 1836</td>
</tr>
<tr>
<td>AMACURDY option</td>
<td>MODEL statement (PANEL), 1821</td>
</tr>
<tr>
<td>ARTEST= option</td>
<td>MODEL statement (PANEL), 1823</td>
</tr>
<tr>
<td>ATOL= option</td>
<td>MODEL statement (PANEL), 1823</td>
</tr>
<tr>
<td>BASE= option</td>
<td>PROC PANEL statement, 1814</td>
</tr>
<tr>
<td>BFN option</td>
<td>MODEL statement (PANEL), 1833</td>
</tr>
<tr>
<td>BIASCORRECTED option</td>
<td>MODEL statement (PANEL), 1824</td>
</tr>
<tr>
<td>BL91 option</td>
<td>MODEL statement (PANEL), 1833</td>
</tr>
<tr>
<td>BL95 option</td>
<td>MODEL statement (PANEL), 1833</td>
</tr>
<tr>
<td>BP option</td>
<td>MODEL statement (PANEL), 1833</td>
</tr>
<tr>
<td>BSY option</td>
<td>MODEL statement (PANEL), 1833</td>
</tr>
<tr>
<td>BTOL= option</td>
<td>MODEL statement (PANEL), 1824</td>
</tr>
<tr>
<td>BTWNG option</td>
<td>MODEL statement (PANEL), 1821</td>
</tr>
<tr>
<td>BW option</td>
<td>MODEL statement (PANEL), 1833</td>
</tr>
<tr>
<td>BY statement</td>
<td>PANEL procedure, 1811</td>
</tr>
<tr>
<td>CDTEST option</td>
<td>MODEL statement (PANEL), 1833</td>
</tr>
<tr>
<td>CLASS statement</td>
<td>PANEL procedure, 1811</td>
</tr>
<tr>
<td>CLUSTER option</td>
<td>Model statement (PANEL), 1824</td>
</tr>
<tr>
<td>COMPARE statement</td>
<td>PANEL procedure, 1811</td>
</tr>
<tr>
<td>CORR option</td>
<td>MODEL statement (PANEL), 1834</td>
</tr>
<tr>
<td>CORRB option</td>
<td>MODEL statement (PANEL), 1834</td>
</tr>
<tr>
<td>CORROUT option</td>
<td>PROC PANEL statement, 1809</td>
</tr>
<tr>
<td>COVB option</td>
<td>MODEL statement (PANEL), 1834</td>
</tr>
<tr>
<td>COVOUT option</td>
<td>MODEL statement (PANEL), 1834</td>
</tr>
<tr>
<td>DASILVA option</td>
<td>MODEL statement (PANEL), 1821</td>
</tr>
<tr>
<td>DATA= option</td>
<td>PROC PANEL statement, 1809</td>
</tr>
<tr>
<td>DLAGS= option</td>
<td>MODEL statement (PANEL), 1824</td>
</tr>
<tr>
<td>DW option</td>
<td>MODEL statement (PANEL), 1834</td>
</tr>
<tr>
<td>DYNDIFF option</td>
<td>MODEL statement (PANEL), 1822</td>
</tr>
<tr>
<td>DYNSYS option</td>
<td>MODEL statement (PANEL), 1822</td>
</tr>
<tr>
<td>FDONE option</td>
<td>MODEL statement (PANEL), 1822</td>
</tr>
<tr>
<td>FDONETIME option</td>
<td>MODEL statement (PANEL), 1822</td>
</tr>
<tr>
<td>FDTWO option</td>
<td>MODEL statement (PANEL), 1822</td>
</tr>
<tr>
<td>FIXONE option</td>
<td>MODEL statement (PANEL), 1822</td>
</tr>
<tr>
<td>FIXONETIME option</td>
<td>MODEL statement (PANEL), 1822</td>
</tr>
<tr>
<td>FIXTWO option</td>
<td>MODEL statement (PANEL), 1822</td>
</tr>
<tr>
<td>FLATDATA statement</td>
<td>PANEL procedure, 1814</td>
</tr>
<tr>
<td>GHM option</td>
<td>MODEL statement (PANEL), 1834</td>
</tr>
<tr>
<td>GINV= option</td>
<td>MODEL statement (PANEL), 1824</td>
</tr>
<tr>
<td>GMM1 option</td>
<td>MODEL statement (PANEL), 1824</td>
</tr>
<tr>
<td>GMM2 option</td>
<td>MODEL statement (PANEL), 1824</td>
</tr>
<tr>
<td>HAC= option</td>
<td>MODEL statement (PANEL), 1825</td>
</tr>
<tr>
<td>HCCME= option</td>
<td>MODEL statement (PANEL), 1826</td>
</tr>
<tr>
<td>HONDA option</td>
<td>MODEL statement (PANEL), 1834</td>
</tr>
<tr>
<td>HONDA2 option</td>
<td>MODEL statement (PANEL), 1834</td>
</tr>
<tr>
<td>HTAYLOR option</td>
<td>MODEL statement (PANEL), 1822</td>
</tr>
</tbody>
</table>
ID statement
   PANEL procedure, 1815
INDID= option
   PROC PANEL statement, 1814
ITGMM option
   MODEL statement (PANEL), 1824
ITPRINT option
   MODEL statement (PANEL), 1834
KEEP= option
   PROC PANEL statement, 1815
KW option
   MODEL statement (PANEL), 1834
LAG statement
   PANEL procedure, 1817
LM option
   TEST statement (PANEL), 1836
LR option
   TEST statement (PANEL), 1836
M= option
   MODEL statement (PANEL), 1823
MAXBAND= option
   INSTRUMENTS statement (PANEL), 1816
MAXITER= option
   MODEL statement (PANEL), 1824
MODEL statement
   PANEL procedure, 1818
MSTAT option
   PROC PANEL statement, 1812
NEWKEYWEST option
   MODEL statement (PANEL), 1826
NOESTIM option
   MODEL statement (PANEL), 1823
NOINT option
   MODEL statement (PANEL), 1823
NOPRINT option
   MODEL statement (PANEL), 1834
OUT= option
   FLATDATA statement (PANEL), 1815
   OUTPUT statement (PANEL), 1835
OUTCORR option
   PROC PANEL statement, 1809
OUTCOV option
   PROC PANEL statement, 1809
OUTTEST= option
   PROC PANEL statement, 1810, 1894
OUTPARM= option
   PROC PANEL statement, 1813
OUTPUT statement
   PANEL procedure, 1835
   PROC PANEL statement, 1893
OUTSTAT= option
   PROC PANEL statement, 1813
OUTTRANS= option
   PROC PANEL statement, 1810, 1895
P= option
   OUTPUT statement (PANEL), 1835
PARKS option
   MODEL statement (PANEL), 1822
PHI option
   MODEL statement (PANEL), 1834
PLOTS= option
   PROC PANEL statement, 1810
POOLED option
   MODEL statement (PANEL), 1822
POOLTEST option
   MODEL statement (PANEL), 1834
PREDICTED= option
   OUTPUT statement (PANEL), 1835
PRINTFIXED option
   MODEL statement (PANEL), 1834
PROC PANEL statement, 1809
PSTAT option
   PROC PANEL statement, 1809
R= option
   OUTPUT statement (PANEL), 1835
RANONE option
   MODEL statement (PANEL), 1822
RANTWO option
   MODEL statement (PANEL), 1822
RESIDUAL= option
   OUTPUT statement (PANEL), 1835
RHO option
   MODEL statement (PANEL), 1834
ROBUST option
   MODEL statement (PANEL), 1824
SINGULAR= option
   MODEL statement (PANEL), 1823
TIME option
   MODEL statement (PANEL), 1824
TSNAME= option
   PROC PANEL statement, 1814
VAR option
   MODEL statement (PANEL), 1834
VCOMP= option
   MODEL statement (PANEL), 1823
WALD option
   TEST statement (PANEL), 1836
WOOLDRIDGE02 option
MODEL statement (PANEL), 1834