THE POWER TO KNOW.

SAS/ETS® 14.3 User's Guide

The HPPANEL Procedure

This document is an individual chapter from *SAS/ETS® 14.3 User's Guide*.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2017. *SAS/ETS® 14.3 User's Guide*. Cary, NC: SAS Institute Inc.

SAS/ETS® 14.3 User's Guide

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

September 2017

Chapter 20
The HPPANEL Procedure

Contents

Overview: HPPANEL Procedure

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

Unlike the PANEL procedure (which can be run only on an individual workstation), the HPPANEL procedure takes advantage of a computing environment that enables it to distribute the optimization task among one or more nodes. Running on one node is called single-machine, and running on more than one node is called distributed mode. In addition, each node (whether in single-machine mode or in distributed mode) can use one or more threads to carry out the optimization on its subset of the data. When several nodes are used and each node uses several threads to carry out its part of the work, the result is a highly parallel computation that provides a dramatic gain in performance.

NOTE: Distributed mode requires SAS High-Performance Econometrics.

You can use the HPPANEL procedure to read and write data in distributed form and perform analyses in distributed mode or in single-machine mode. For more information about how to affect the execution mode of SAS high-performance analytical procedures, see the section “Processing Modes” (Chapter 2, *SAS/ETS User’s Guide: High-Performance Procedures*).

The HPPANEL procedure is specifically designed to operate in the high-performance distributed mode. By default, PROC HPPANEL performs computations in multiple threads.

The panel data models can be grouped into several categories that depend on the structure of the error term. The HPPANEL procedure uses the following error structures and the corresponding methods to analyze data:

- one-way and two-way models
- fixed-effects and random-effects models

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

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

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

Getting Started: HPPANEL Procedure

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

```
data greene;
   input firm year production cost @@;
datalines;
1 1955   5.36598   1.14867  1 1960   6.03787   1.45185
1 1965   6.37673   1.52257  1 1970   6.93245   1.76627
2 1955   6.54535   1.35041  2 1960   6.69827   1.71109
2 1965   7.40245   2.09519  2 1970   7.82644   2.39480
3 1955   8.07153   2.94628  3 1960   8.47679   3.25967

   ... more lines ...
```

You decide to fit the following model to the data,

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

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

If you assume that the time and cross-sectional effects are random, four possible estimators are left for the variance components. The following statements choose the Fuller-Battese method to fit this model:

```
proc hppanel data=greene;
   model cost = production / rantwo vcomp = fb;
   id firm year;
   performance nodes=0 nthreads=2;
run;
```

The output of the HPPANEL procedure is shown in Output 20.1.

Figure 20.1 Two-Way Random Effects Results

The HPPANEL Procedure

Model Information	
Data Source	GREENE
Response Variable	cost
Model	RANTWO
Variance Component	FULLER

Fit Statistics	
Sum of Squared Error	0.348082
Degrees of Freedom	22
Mean Squared Error	0.015822
Root Mean Squared Error	0.125785
R-Square	0.813624

Variance Component Estimates	
Variance Component for Cross Sections	0.0469
Variance Component for Time Series	0.00906
Variance Component for Error	0.00875

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-2.99992	0.64778	-4.63	<.0001
production	1	0.74660	0.07618	9.80	<.0001

Printed first is the model description, which reports the method used for estimation and the method used for estimating error components. Printed next is the fit statistics table, and then the variance components estimates. Finally, the table of regression parameter estimates shows the estimates, standard errors, and *t* tests.

Syntax: HPPANEL Procedure

The following statements are available in the HPPANEL procedure:

PROC HPPANEL *options* ;
    ID *cross-section-id time-series-id* ;
    MODEL *response = regressors* < */options* > ;
    RESTRICT *equation1* < *,equation2. . .* > ;
    TEST *equation* < *,equation2. . .* >< */ options* > ;
    OUTPUT OUT=*SAS-data-set* < *output-options* > ;
    PERFORMANCE < *performance-options* > ;

The ID and MODEL statements are required.

The following sections provide a functional summary of statements and options, describe the PROC HPPANEL statement, and then describe the other statements in alphabetical order.

Functional Summary

Table 20.1 summarizes the statements and options that you can use in the HPPANEL procedure.

Table 20.1 Functional Summary

Description	Statement	Option
Data Set Options		
Includes correlations in the OUTEST= data set	PROC HPPANEL	CORROUT
Includes covariances in the OUTEST= data set	PROC HPPANEL	COVOUT
Specifies the input data set	PROC HPPANEL	DATA=
Specifies the name of an output SAS data set	OUTPUT	OUT=
Writes parameter estimates to an output data set	PROC HPPANEL	OUTEST=
Variable Role Options		
Specifies the cross-sectional and time ID variables	ID	
Printing Control Options		
Prints correlations of the estimates	PROC HPPANEL	CORRB
Prints covariances of the estimates	PROC HPPANEL	COVB
Suppresses printed output	PROC HPPANEL	NOPRINT
Prints fixed effects	MODEL	PRINTFIXED
Performs tests of linear hypotheses	TEST	
Model Estimation Options		
Estimates the between-groups model	MODEL	BTWNG
Estimates the between-time-periods model	MODEL	BTWNT
Estimates the one-way fixed-effects model	MODEL	FIXONE

Table 20.1 *continued*

Description	Statement	Option
Estimates the one-way fixed-effects model with respect to time	MODEL	FIXONETIME
Estimates the two-way fixed-effects model	MODEL	FIXTWO
Suppresses the intercept term	MODEL	NOINT
Estimates the pooled regression model	MODEL	POOLED
Estimates the one-way random-effects model	MODEL	RANONE
Estimates the two-way random-effects model	MODEL	RANTWO
Specifies the method for the variance components estimator	MODEL	VCOMP=
Specifies linear equality restrictions on the parameters	RESTRICT	
Specifies which tests to perform	TEST	WALD, LM, LR

PROC HPPANEL Statement

PROC HPPANEL *options* **;**

The HPPANEL statement invokes the HPPANEL procedure.

You can specify the following *options*:

DATA=*SAS-data-set*
names the input data set. Only one observation is allowed for each cross section and time period. If you omit the DATA= option, PROC HPPANEL uses the most recently created SAS data set.

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

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

NOPRINT
suppresses the normal printed output.

OUTEST=*SAS-data-set*
names an output data set to contain the parameter estimates. When the OUTEST= option is not specified, the OUTEST= data set is not created. For more information about the structure of the OUTEST= data set, see the section "OUTEST= Data Set" on page 1134.

OUTCOV
COVOUT
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 1134.

OUTCORR

CORROUT

writes the correlation matrix of the parameter estimates to the OUTEST= data set. For more information, see the section "OUTEST= Data Set" on page 1134.

In addition, you can specify any of the following MODEL statement options in the PROC HPPANEL statement: FIXONE, FIXONETIME, FIXTWO, RANONE, RANTWO, NOINT, PRINTFIXED, and VCOMP=. Specifying these options in the PROC HPPANEL statement is equivalent to specifying them in the MODEL statement. For a complete description of each of these options, see the section "MODEL Statement" on page 1121.

ID Statement

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

The ID statement specifies variables in the input data set that identify the cross section and the time period for each observation. The ID statement is required. Unlike the PANEL procedure, the HPPANEL procedure does not require the data set to be sorted.

MODEL Statement

MODEL *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 is regressed on the independent variables (*regressors*). You can specify only one MODEL statement and only one *response*.

You specify the error structure and estimation technique by including one of the following *options* after a slash (/):

BTWNG

estimates the between-groups model.

BTWNT

estimates the between-time-periods model.

FIXONE

estimates a one-way fixed-effects model, which corresponds to cross-sectional effects.

FIXONETIME

estimates a one-way fixed-effects model, which corresponds to time effects.

FIXTWO

estimates a two-way fixed-effects model.

POOLED

estimates the pooled regression model.

RANONE
estimates a one-way random-effects model.

RANTWO
estimates a two-way random-effects model.

By default, a FIXONE estimation is performed.

You can also specify the following *options* after the slash:

NOINT
suppresses the intercept parameter from the model.

PRINTFIXED
prints the fixed effects.

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

FB	requests the Fuller-Battese estimator.
WK	requests the Wansbeek-Kapteyn estimator.
WH	requests the Wallace-Hussain estimator.
NERLOVE	requests the Nerlove estimator.

By default, VCOMP=WK for both balanced and unbalanced data.

OUTPUT Statement

OUTPUT OUT=*SAS-data-set* < *output-options* > **;**

The OUTPUT statement creates a new SAS data set to contain variables that are specified by the COPYVAR option, the cross-sectional ID (_CSID_), and the time period (_TSID_). This data set also contains the predicted value and the residual if they are specified by *output-options*. When the response values are missing for the observation, all output estimates except the residual are still computed as long as none of the explanatory variables are missing. You can specify only one OUTPUT statement.

You must specify the OUT= option:

OUT=*SAS-data-set*
names the output data set.

You can specify one or more of the following *output-options*:

COPYVAR=(*SAS-variable-names***)**
COPYVARS=(*SAS-variable-names***)**
adds SAS variables to the output data set.

PREDICTED
outputs estimates of predicted dependent variables.

RESIDUAL
outputs estimates of residuals.

PERFORMANCE Statement

PERFORMANCE < *performance-options* > **;**

The PERFORMANCE statement specifies *performance-options* to control the multithreaded and distributed computing environment and requests detailed performance results of the HPPANEL procedure. You can also use the PERFORMANCE statement to control whether the HPPANEL procedure executes in single-machine or distributed mode. You can specify the following *performance-options*:

DETAILS
requests a table that shows a timing breakdown of the procedure steps.

NODES=*n*
specifies the number of nodes in the distributed computing environment, provided that the data are not processed alongside the database.

NTHREADS=*n*
specifies the number of threads for analytic computations and overrides the SAS system option THREADS | NOTHREADS. If you do not specify the NTHREADS= option, PROC HPPANEL creates one thread per CPU for the analytic computations.

The PERFORMANCE statement is documented further in the section “PERFORMANCE Statement” (Chapter 2, *SAS/ETS User’s Guide: High-Performance Procedures*).

RESTRICT Statement

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

The RESTRICT statement specifies linear equality restrictions on the parameters in the MODEL statement. There can be as many unique restrictions as the number of parameters in the MODEL statement. Multiple RESTRICT statements are understood as joint restrictions on the model’s parameters.

Currently, PROC HPPANEL only supports linear equality restrictions. Restriction expressions can be composed only of algebraic operations that involve the addition symbol (+), subtraction symbol (–), and multiplication symbol (*).

The following statements illustrate the use of the RESTRICT statement:

```
proc hppanel;
   id csid tsid;
   model y = x1 x2 x3;
   restrict x1 = 0, x2 * .5 + 2 * x3= 0;
   restrict  x2 = 0, intercept = 0;
run;
```

A RESTRICT statement cannot include a division sign in its formulation. As in the preceding example, you can obtain restrictions on the intercept by using the keyword INTERCEPT.

TEST Statement

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

The TEST statement performs Wald, Lagrange multiplier, and likelihood ratio tests of linear hypotheses about the regression parameters in the MODEL statement. Each *equation* specifies a linear hypothesis to be tested. Currently, only linear equality restrictions and tests are permitted in PROC HPPANEL. Test expressions can be composed only of algebraic operations that involve the addition symbol (+), subtraction symbol (–), and multiplication symbol (*). 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. In the equality restrictions, you can use the keyword INTERCEPT to refer to the coefficient of the intercept.

You can specify the following *options* after the slash (/):

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

WALD
: specifies the Wald test.

LM
: specifies the Lagrange multiplier test.

LR
: specifies the likelihood ratio test.

By default, the Wald test is performed.

The following statements illustrate the use of the TEST statement:

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

The first test investigates the joint hypothesis that

$$\beta_1 = 0$$

and

$$0.5\beta_2 + 2\beta_3 = 0$$

Details: HPPANEL Procedure

Specifying the Input Data

The HPPANEL procedure is similar to other regression procedures in SAS. Suppose you want to regress the variable Y on regressors X1 and X2. Cross sections are identified by the variable State, and time periods are identified by the variable Date. Unlike the PANEL procedure, the HPPANEL procedure does not require the data set to be sorted. To invoke the HPPANEL procedure, you must specify the cross section and time series variables in an ID statement. The following statements show the correct syntax:

```
proc hppanel data=a;
   id state date;
   model y = x1 x2;
   performance nodes=2 nthreads=4;
run;
```

Specifying the Regression Model

The MODEL statement in PROC HPPANEL is specified like the MODEL statement in other SAS regression procedures: the dependent variable is listed first, followed by an equal sign, followed by the list of regressor variables, as shown in the following statements:

```
proc hppanel data=a;
   id state date;
   model y = x1 x2;
   performance nodes=2 nthreads=4;
run;
```

Specifying the Number of Nodes and Number of Threads

The PERFORMANCE statement in PROC HPPANEL is specified like the PERFORMANCE statement in other SAS high-performance procedures. The following statements execute the model in the distributed computing environment with two threads and four nodes:

```
proc hppanel data=a;
   id state date;
   model y = x1 x2;
   performance nodes=2 nthreads=4;
run;
```

The major advantage of using PROC HPPANEL is that you can incorporate a model for the structure of the random errors. It is important to consider what type of error structure model is appropriate for your data and to specify the corresponding option in the MODEL statement.

The error structure options supported by the HPPANEL procedure are FIXONE, FIXONETIME, FIXTWO, RANONE, and RANTWO. For more information about these methods and the error structures they assume, see the following sections. The following statements fit a Fuller-Battese one-way random-effects model:

```
proc hppanel data=a;
   id state date;
   model y = x1 x2 / ranone vcomp=fb;
   performance nodes=0 nthreads=1;
run;
```

To aid in model specification within this class of models, PROC HPPANEL provides one specification test statistic, the Hausman *m* statistic, which provides information about the appropriateness of the random-effects specification. The *m* statistic is based on the idea that, under the null hypothesis of no correlation between the effects variables and the regressors, ordinary least squares (OLS) and generalized least squares (GLS) are consistent. However, OLS is inefficient. Hence, a test can be based on the result that the covariance between an efficient estimator and its difference from an inefficient estimator is 0. Rejection of the null hypothesis might suggest that the fixed-effects model is more appropriate.

The HPPANEL procedure also provides the Buse R-square measure. This number is interpreted as a measure of the proportion of the transformed sum of squares of the dependent variable that is attributable to the influence of the independent variables. For OLS estimation, the Buse R-square measure is equivalent to the usual R-square measure.

Unbalanced Data

The HPPANEL procedure can process data that have different numbers of time series observations across different cross sections. The missing time series observations are recognized by the absence of time series ID variable values in some of the cross sections in the input data set. Moreover, if an observation that has a particular time series ID value and cross-sectional ID value is present in the input data set but one or more of the model variables are missing, that time series point is treated as missing for that cross section.

One-Way Fixed-Effects Model

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

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

where the γ_i are nonrandom parameters to be estimated.

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

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

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

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

After the slope estimates have been calculated, the estimation of an intercept or the cross-sectional fixed effects is handled as follows. First, you obtain the cross-sectional effects:

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

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

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

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

The within-model sum of squared errors is

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

The estimated error variance can be written as

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

Alternatively, an equivalent way to express the error variance is

$$\hat{\sigma}_\epsilon^2 = \tilde{\mathbf{u}}^{'} \mathbf{Q}_0 \tilde{\mathbf{u}}/(M - N - (K-1))$$

where the residuals $\tilde{\mathbf{u}}$ are given by $\tilde{\mathbf{u}} = (\mathbf{I}_M - \mathbf{j}_M \mathbf{j'}_M/M)(\mathbf{y} - \mathbf{X}_s \tilde{\beta}_s)$ if there is an intercept and by $\tilde{\mathbf{u}} = (\mathbf{y} - \mathbf{X}_s \tilde{\beta}_s)$ if there is not. The drawback is that the formula changes (but the results do not) with the inclusion of a constant.

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

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

The covariance of the dummy variables and the dummy variables with the $\tilde{\beta}_s$ depends on whether the intercept is included in the model. For more information, see the section "One-Way Fixed-Effects Model (FIXONE and FIXONETIME Options)" on page 1840.

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

Two-Way Fixed-Effects Model

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

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

where the γ_i and α_t are nonrandom parameters to be estimated.

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

Balanced Panels

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

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

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

where the symbols are as follows:

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

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

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

The following calculations of cross-sectional dummy variables, time dummy variables, and intercepts are similar to how they are calculated in the one-way model:

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

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

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

Use the superscript C and T to denote the cross-sectional dummy variables and time dummy variables, respectively. Under the NOINT option, the following equations produce the dummy variables:

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

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

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

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

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

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

The sum of squared errors is

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

The estimated error variance is

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

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

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

For information about the covariance matrix that is related to dummy variables, see the section "Two-Way Random-Effects Model (RANTWO Option)" on page 1846.

Unbalanced Panels

Let $\mathbf{X}_*$ and $\mathbf{y}_*$ be the independent and dependent variables, respectively, that are arranged by time and by cross section within each time period. (Note that the input data set that the PANEL procedure uses must be sorted by cross section and then by time within each cross section.) Let M_t be the number of cross sections that are observed in year *t*, and let $\sum_t M_t = M$. Let $\mathbf{D}_t$ be the $M_t \times N$ matrix that is obtained from the $N \times N$ identity matrix from which rows that correspond to cross sections that are not observed at time *t* have been omitted. Consider

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

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

Let

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

The estimate of the regression slope coefficients is given by

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

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

The estimator of the error variance is

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

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

The actual implementation is quite different from the theory. For more information, see the section "Two-Way Fixed-Effects Model (FIXTWO Option)" on page 1841.

One-Way Random-Effects Model

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

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

Let $\mathbf{Z}_0 = \text{diag}(\mathbf{J}_{T_i})$, $\mathbf{P}_0 = \text{diag}(\bar{\mathbf{J}}_{T_i})$, and $\mathbf{Q}_0 = \text{diag}(\mathbf{E}_{T_i})$, with $\bar{\mathbf{J}}_{T_i} = \mathbf{J}_{T_i}/T_i$ and $\mathbf{E}_{T_i} = \mathbf{I}_{T_i} - \bar{\mathbf{J}}_{T_i}$. Define $\tilde{\mathbf{X}}_s = \mathbf{Q}_0\mathbf{X}_s$. Also define $\tilde{\mathbf{y}} = \mathbf{Q}_0\mathbf{y}$ and $\mathbf{J}$ as a vector of 1s whose length is T_i.

In the one-way model, estimation proceeds in a two-step fashion. First, you obtain estimates of the variance of the σ_ϵ^2 and σ_ν^2. There are multiple ways to derive these estimates; PROC HPPANEL provides four options. For more information, see the section "One-Way Random-Effects Model (RANONE Option)" on page 1843.

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

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

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

where T_i is the ith cross section's time observations.

Taking the θ_i, you form the partial deviations,

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

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

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

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

Two-Way Random-Effects Model

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

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

As it does for the one-way random-effects model, the HPPANEL procedure provides four options for variance component estimators. However, unbalanced panels present some special concerns that do not occur for one-way random-effects models.

Let $\mathbf{X}_*$ and $\mathbf{y}_*$ be the independent and dependent variables that are arranged by time and by cross section within each time period. (Note that the input data set that the PANEL procedure uses must be sorted by cross section and then by time within each cross section.) Let M_t be the number of cross sections that are observed in time t, and let $\sum_t M_t = M$. Let $\mathbf{D}_t$ be the $M_t \times N$ matrix that is obtained from the $N \times N$ identity matrix from which rows that correspond to cross sections that are not observed at time t have been omitted. Consider

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

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

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

For notational ease, let

$$\Delta_N = \mathbf{Z}_1^{'}\mathbf{Z}_1$$

$$\Delta_T = \mathbf{Z}_2^{'}\mathbf{Z}_2$$

$$\mathbf{A} = \mathbf{Z}_2^{'}\mathbf{Z}_1$$

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

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

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

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

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

PROC HPPANEL provides four methods to estimate the variance components. For more information, see the section “Two-Way Random-Effects Model (RANTWO Option)” on page 1846.

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

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

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

The θ estimates are obtained from

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

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

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

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

The case of an unbalanced panel is somewhat more complicated. Wansbeek and Kapteyn show that the inverse of Ω can be written as

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

with the following:

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

By using the inverse of the covariance matrix of the error, it becomes possible to complete GLS on the unbalanced panel.

Between Estimators

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

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

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

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

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

Pooled Estimator

The pooled estimator is simply linear regression that is run on all the data, without regard to cross section or time:

$$y_{it} = \mathbf{x}_{it}\beta^{P} + u_{it}$$

The error is assumed to be normally distributed with mean zero and a constant variance.

Linear Hypothesis Testing

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

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

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

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

where

- $\hat{\mathbf{u}}_{*}$is the residual vector from the restricted regression
- $\hat{\mathbf{u}}$ is the residual vector from the unrestricted regression
- J is the number of restrictions

- $M-K$ are the degrees of freedom, M is the number of observations, and K is the number of parameters in the model

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

The Wald test statistic is

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

The likelihood ratio is

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

The Lagrange multiplier test statistic is

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

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

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

Specification Tests

The HPPANEL procedure outputs one specification test for random effects: the Hausman (1978) specification test (m statistic) can be used to test hypotheses in terms of bias or inconsistency of an estimator. This test was also proposed by Wu (1973) and further extended in Hausman and Taylor (1982). Hausman's m statistic is as follows.

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

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

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

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

OUTPUT OUT= Data Set

PROC HPPANEL writes the initial data of the estimated model, predicted values, and residuals to an output data set when the OUT= option is specified in the OUTPUT statement. The OUT= data set contains the following variables:

CSID is the value of the cross section ID. The variable name is the one specified in the id statement.

TSID is the value of the time period in the dynamic model. The variable name is the one specified in the id statement.

Regressors are the values of regressor variables that are specified in the COPYVAR option.

Pred is the predicted value of dependent variable. This column is output only if the PRED option is specified.

Resid is the residual from the regression. This column is output only if the RESIDUAL option is specified.

OUTEST= Data Set

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

METHOD is a character variable that identifies the estimation method.

TYPE is a character variable that identifies the type of observation. Values of the _TYPE_ variable are CORRB, COVB, CSPARMS, STD, and the type of model estimated. The CORRB observation contains correlations of the parameter estimates; the COVB observation contains covariances of the parameter estimates; the STD observation indicates the row of standard deviations of the corresponding coefficients; and the type of model estimated observation contains the parameter estimates.

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

DEPVAR is a character variable that contains the name of the response variable.

MSE is the mean square error of the transformed model.

VARCS is the variance component estimate due to cross sections. The _VARCS_ variable is included in the OUTEST= data set when the RANONE option is specified in the MODEL or PROC HPPANEL statement.

VARTS is the variance component estimate due to time series. The _VARTS_ variable is included in the OUTEST= data set when the RANTWO option is specified in the MODEL or PROC HPPANEL statement.

VARERR is the variance component estimate due to error. The _VARERR_ variable is included in the OUTEST= data set when the RANONE or RANTWO option is specified in the MODEL or PROC HPPANEL statement.

Intercept is the intercept parameter estimate. (The intercept is missing for models when the NOINT option is specified in the MODEL statement.)

Regressors are the regressor variables that are specified in the MODEL statement. The regressor variables in the OUTEST= data set contain the corresponding parameter estimates, and the corresponding covariance or correlation matrix elements for _TYPE_=COVB and _TYPE_=CORRB observations.

Printed Output

The printed output from PROC HPPANEL includes the following:

- the model information, which includes the data source, the dependent variable name, the estimation method used, and for random-effects model analysis, the variance component estimation method.
- the number of observations
- the fit statistics, which include the sum of squared error (SSE), the degree of freedom for error (DFE), the mean square error (MSE), the root mean square error (RMSE), and the R-square
- the error components estimates for random-effects model
- the Hausman test statistics, which include the degree of freedom (DF), the test statistics, and the *p*-value.
- the regression parameter estimates and analysis, which include for each regressor 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 HPPANEL prints the following:

- the covariance and correlation of the resulting regression parameter estimates
- the WALD, LR, and LM test statistics for linear equality restrictions that are specified in the TEST statements
- the timing breakdown of the procedure steps

ODS Table Names

PROC HPPANEL assigns a name to each table it creates. You can use these names to refer to the table when you use the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in Table 20.2.

Table 20.2 ODS Tables Produced in PROC HPPANEL

ODS Table Name	Description	Option
ODS Tables Created by the MODEL Statement		
ModelInfo	Model information	Default
PerformanceInfo	Performance information	Default
Nobs	Number of observations	Default
FitStatistics	Fit statistics	Default
ParameterEstimates	Parameter estimates	Default
CovB	Covariance of parameter estimates	COVB
CorrB	Correlations of parameter estimates	CORRB
RandomEffectsTest	Hausman test for random effects	RANONE, RANTWO
ODS Tables Created by the TEST Statement		
TestResults	Test results	
ODS Tables Created by the PERFORMANCE Statement		
Timing	Timing Table	

Example: HPPANEL Procedure

Example 20.1: One-Way Random-Effects High-Performance Model

This example shows the use of the one-way random-effects model that is available in the HPPANEL procedure; the example emphasizes processing a large data set and the performance improvements that are achieved by executing in a high-performance distributed environment.

The following DATA step generates five million observations from one-way panel data that includes 50,000 cross sections and 100 time periods:

```
data hppan_ex01 (keep = cs ts y x1-x10);
   retain seed 55371;
   array x[10];
   label y  = 'Dependent Variable';
   do cs = 1 to 50000;
      dummy = 10 * rannor(seed);
      do ts = 1 to 100;
      /*- generate regressors and compute the structural */
      /*- part of the dependent variable                 */
         y = 5;
         do k = 1 to 10;
            x[k] = -1 + 2 * ranuni(seed);
            y = y + x[k] * k;
         end;

         /*- add an error term, such that e - N(0,100)   -------*/
         y = y + 10 * rannor(seed);
         /*- add a random effect, such that v - N(0,100) -------*/
         y = y + dummy;
         output;
      end;
   end;
run;
```

The estimation is executed in distributed mode on a grid with ten nodes, with one thread per node. To run the following statements successfully, you need to set the macro variables GRIDHOST and GRIDINSTALLLOC to resolve to appropriate values, or you can replace the references to the macro variables in the example with the appropriate values.

```
%let GRIDHOST      = <<your grid host>>;
%let GRIDINSTALLOC = <<your grid install location>>;

option set = GRIDHOST       = "&GRIDHOST";
option set = GRIDINSTALLLOC = "&GRIDINSTALLLOC";

proc hppanel data=hppan_ex01;
   id  cs ts;
   model y = x1-x10 / ranone;
   performance nodes = 10 threads = 1 details
               host="&GRIDHOST" install="&GRIDINSTALLLOC";
run;
```

In Output 20.1.1, the “Performance Information” table shows that the model was estimated on the grid that is defined in the macro variable named GRIDHOST in a distributed environment with ten nodes, and one thread per node. The grid installation location is defined in the macro variable named GRIDINSTALLLOC.

Output 20.1.1 Grid Information with Ten Nodes and One Thread per Node

Performance Information	
Host Node	<< your grid host >>
Install Location	<< your grid install location >>
Execution Mode	Distributed
Number of Compute Nodes	10
Number of Threads per Node	1

Output 20.1.2 shows the results for the one-way random-effects model. The "Model Information" table shows detailed information about the model. The "Number of Observations" table indicates that all five million observations were used to fit the model. All parameter estimates in the "Parameter Estimates" table are highly significant and correspond to the theoretical values that were set for them during the data generating process. In the "Procedure Task Timing" table, you can see that for five million observations, computing the moments took 101.53 seconds, and the time taken for cross-product accumulation was negligible.

Output 20.1.2 One-Way Random-Effects Model

Model Information	
Data Source	HPPAN_EX01
Response Variable	y
Model	RANONE
Variance Component	WANSBEEK

Number of Observations	
Number of Observations Read	5000000
Number of Observations Used	5000000
Number of Cross Sections	50000
Number of Time Series	100

Fit Statistics	
Sum of Squared Error	4.9976E8
Degrees of Freedom	4999989
Mean Squared Error	99.952
Root Mean Squared Error	9.9976
R-Square	0.559771

Variance Component Estimates	
Variance Component for Cross Sections	99.9117
Variance Component for Error	99.9520

Hausman Test for Random Effects

Coefficients	DF	m Value	Pr > m
10	10	14.04	0.1713

Output 20.1.2 *continued*

Parameter Estimates

Parameter	DF	Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	4.96955	0.04492	110.62	<.0001
x1	1	1.00902	0.00778	129.69	<.0001
x2	1	1.99743	0.00778	256.66	<.0001
x3	1	3.00116	0.00778	385.64	<.0001
x4	1	3.99847	0.00778	513.68	<.0001
x5	1	4.99497	0.00778	641.81	<.0001
x6	1	6.01034	0.00778	772.12	<.0001
x7	1	6.99770	0.00778	899.39	<.0001
x8	1	7.98897	0.00778	1026.61	<.0001
x9	1	9.00692	0.00778	1157.12	<.0001
x10	1	10.00563	0.00778	1285.47	<.0001

Procedure Task Timing

Task	Seconds	Percent
Data Read and Variable Levelization	0.29	5.31%
Communication to Client	0.00	0.00%
Computing Moments	4.34	78.92%
Cross-Product Accumulation	0.87	15.77%

For comparison, you now fit a pooled regression estimation on the same data, again using a grid of 10 nodes with one thread each. The following SAS statements perform the estimation on the grid:

```
proc hppanel data=hppan_ex01;
   id  cs ts;
   model y = x1-x10 / pooled;
   performance nodes = 10 threads = 1 details
               host="&GRIDHOST" install="&GRIDINSTALLLOC";
run;
```

Based on Output 20.1.3, you find that the parameter estimates are similar to those from the random-effects estimator. You also find that the timings are similar, indicating that the bulk of the computational effort is due to tasks common to both random-effects estimation and standard OLS regression. In both cases, estimation is dominated by the calculation of sums of squares and other moment terms, over the whole data set.

Output 20.1.3 Pooled Regression Model

The HPPANEL Procedure

Model Information	
Data Source	HPPAN_EX01
Response Variable	y
Model	POOLED

Output 20.1.3 *continued*

Parameter Estimates					
Parameter	**DF**	**Estimate**	**Standard Error**	**t Value**	**Pr > |t|**
Intercept	1	4.96957	0.00632	786.03	<.0001
x1	1	1.01251	0.01095	92.49	<.0001
x2	1	1.98374	0.01095	181.17	<.0001
x3	1	3.00294	0.01095	274.23	<.0001
x4	1	3.99649	0.01095	364.90	<.0001
x5	1	5.00187	0.01095	456.77	<.0001
x6	1	5.99952	0.01095	547.77	<.0001
x7	1	7.00478	0.01095	639.88	<.0001
x8	1	7.97232	0.01095	728.13	<.0001
x9	1	9.01244	0.01095	822.90	<.0001
x10	1	10.01578	0.01095	914.52	<.0001

Procedure Task Timing		
Task	**Seconds**	**Percent**
Data Read and Variable Levelization	0.28	5.74%
Communication to Client	0.00	0.00%
Computing Moments	4.29	87.33%
Cross-Product Accumulation	0.34	6.92%

References

Davidson, R., and MacKinnon, J. G. (1993). *Estimation and Inference in Econometrics*. New York: Oxford University Press.

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

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

Greene, W. H. (2000). *Econometric Analysis*. 4th ed. Upper Saddle River, NJ: Prentice-Hall.

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

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

Nerlove, M. (1971). "Further Evidence on the Estimation of Dynamic Relations from a Time Series of Cross Sections." *Econometrica* 39:359–382.

Wallace, T., and Hussain, A. (1969). "The Use of Error Components Model in Combining Cross Section with Time Series Data." *Econometrica* 37:55–72.

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

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

Subject Index

Syntax Index