

xtpcse — Linear regression with panel-corrected standard errors[Description](#)[Quick start](#)[Menu](#)[Syntax](#)[Options](#)[Remarks and examples](#)[Stored results](#)[Methods and formulas](#)[Acknowledgments](#)[References](#)[Also see](#)

Description

`xtpcse` calculates panel-corrected standard error (PCSE) estimates for linear cross-sectional time-series models where the parameters are estimated by either OLS or Prais–Winsten regression. When computing the standard errors and the variance–covariance estimates, `xtpcse` assumes that the disturbances are, by default, heteroskedastic and contemporaneously correlated across panels.

See [\[XT\] xtglm](#) for the generalized least-squares estimator for these models.

Quick start

Linear regression of `y` on `x1` and `x2` with panel-corrected standard errors and assuming no within-panel autocorrelation using `xtset` data

```
xtpcse y x1 x2
```

Same as above, but specify a common first-order autocorrelation within panels

```
xtpcse y x1 x2, correlation(ar1)
```

Within-panel heteroskedastic errors but no contemporaneous correlation between panels

```
xtpcse y x1 x2, hetonly
```

Let autocorrelation structure be panel-specific estimated by time-series methods

```
xtpcse y x1 x2, correlation(psar1) rhotype(tscorr)
```

Menu

Statistics > Longitudinal/panel data > Contemporaneous correlation > Regression with panel-corrected standard errors (PCSE)

Syntax

```
xtpcse depvar [indepvars] [if] [in] [weight] [, options]
```

| <i>options</i> | Description |
|--|--|
| Model | |
| <code>noconstant</code> | suppress constant term |
| <code>correlation(<u>i</u>ndependent)</code> | use independent autocorrelation structure |
| <code>correlation(ar1)</code> | use AR1 autocorrelation structure |
| <code>correlation(<u>p</u>sar1)</code> | use panel-specific AR1 autocorrelation structure |
| <code><u>r</u>hotype(<i>calc</i>)</code> | specify method to compute autocorrelation parameter; seldom used |
| <code>np1</code> | weight panel-specific autocorrelations by panel sizes |
| <code><u>h</u>etonly</code> | assume panel-level heteroskedastic errors |
| <code><u>i</u>ndependent</code> | assume independent errors across panels |
| by/if/in | |
| <code><u>c</u>asewise</code> | include only observations with complete cases |
| <code><u>p</u>airwise</code> | include all available observations with nonmissing pairs |
| SE | |
| <code><u>n</u>mk</code> | normalize standard errors by $N - k$ instead of N |
| Reporting | |
| <code><u>l</u>evel(#)</code> | set confidence level; default is <code>level(95)</code> |
| <code><u>d</u>etail</code> | report list of gaps in time series |
| <code><i>display_options</i></code> | control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling |
| <code><u>c</u>oeflegend</code> | display legend instead of statistics |

A panel variable and a time variable must be specified; use `xtset`; see [XT] `xtset`.

indepvars may contain factor variables; see [U] 11.4.3 **Factor variables**.

depvar and *indepvars* may contain time-series operators; see [U] 11.4.4 **Time-series varlists**.

by, collect, and statsby are allowed; see [U] 11.1.10 **Prefix commands**.

iweights and aweights are allowed; see [U] 11.1.6 **weight**.

coeflegend does not appear in the dialog box.

See [U] 20 **Estimation and postestimation commands** for more capabilities of estimation commands.

Options

Model

`noconstant`; see [R] **Estimation options**.

`correlation(corr)` specifies the form of assumed autocorrelation within panels.

`correlation(independent)`, the default, specifies that there is no autocorrelation.

`correlation(ar1)` specifies that, within panels, there is first-order autocorrelation AR(1) and that the coefficient of the AR(1) process is common to all the panels.

`correlation(psar1)` specifies that, within panels, there is first-order autocorrelation and that the coefficient of the AR(1) process is specific to each panel. `psar1` stands for panel-specific AR(1).

`rhotype(calc)` specifies the method to be used to calculate the autocorrelation parameter. Allowed strings for `calc` are

| | |
|----------------------|---|
| <code>regress</code> | regression using lags; the default |
| <code>freg</code> | regression using leads |
| <code>tscorr</code> | time-series autocorrelation calculation |
| <code>dw</code> | Durbin–Watson calculation |

All the above methods are consistent and asymptotically equivalent; this is a rarely used option.

`np1` specifies that the panel-specific autocorrelations be weighted by T_i rather than by the default $T_i - 1$ when estimating a common ρ for all panels, where T_i is the number of observations in panel i . This option has an effect only when panels are unbalanced and the `correlation(ar1)` option is specified.

`hetonly` and `independent` specify alternative forms for the assumed covariance of the disturbances across the panels. If neither is specified, the disturbances are assumed to be heteroskedastic (each panel has its own variance) and contemporaneously correlated across the panels (each pair of panels has its own covariance). This is the standard PCSE model.

`hetonly` specifies that the disturbances are assumed to be panel-level heteroskedastic only with no contemporaneous correlation across panels.

`independent` specifies that the disturbances are assumed to be independent across panels; that is, there is one disturbance variance common to all observations.

by/if/in

`casewise` and `pairwise` specify how missing observations in unbalanced panels are to be treated when estimating the interpanel covariance matrix of the disturbances. The default is `casewise` selection.

`casewise` specifies that the entire covariance matrix be computed only on the observations (periods) that are available for all panels. If an observation has missing data, all observations of that period are excluded when estimating the covariance matrix of disturbances. Specifying `casewise` ensures that the estimated covariance matrix will be of full rank and will be positive definite.

`pairwise` specifies that, for each element in the covariance matrix, all available observations (periods) that are common to the two panels contributing to the covariance be used to compute the covariance.

The `casewise` and `pairwise` options have an effect only when the panels are unbalanced and neither `hetonly` nor `independent` is specified.

SE

`nmk` specifies that standard errors be normalized by $N - k$, where k is the number of parameters estimated, rather than N , the number of observations. Different authors have used one or the other normalization. [Greene \(2018, 313\)](#) remarks that whether a degree-of-freedom correction improves the small-sample properties is an open question.

Reporting

`level(#)`; see [\[R\] Estimation options](#).

`detail` specifies that a detailed list of any gaps in the series be reported.

display_options: `noci`, `nopvalues`, `noomitted`, `vsquish`, `noemptycells`, `baselevels`, `allbaselevels`, `novlabel`, `fvwrap(#)`, `fvwrapon(style)`, `cformat(%fmt)`, `pformat(%fmt)`, `sformat(%fmt)`, and `nolstretch`; see [R] [Estimation options](#).

The following option is available with `xtpcse` but is not shown in the dialog box:

`coeflegend`; see [R] [Estimation options](#).

Remarks and examples

[stata.com](http://www.stata.com)

`xtpcse` is an alternative to feasible generalized least squares (FGLS)—see [XT] [xtgls](#)—for fitting linear cross-sectional time-series models when the disturbances are not assumed to be independent and identically distributed (i.i.d.). Instead, the disturbances are assumed to be either heteroskedastic across panels or heteroskedastic and contemporaneously correlated across panels. The disturbances may also be assumed to be autocorrelated within panel, and the autocorrelation parameter may be constant across panels or different for each panel.

We can write such models as

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \epsilon_{it}$$

where $i = 1, \dots, m$ is the number of units (or panels); $t = 1, \dots, T_i$; T_i is the number of periods in panel i ; and ϵ_{it} is a disturbance that may be autocorrelated along t or contemporaneously correlated across i .

This model can also be written panel by panel as

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_m \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_m \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \vdots \\ \boldsymbol{\epsilon}_m \end{bmatrix}$$

For a model with heteroskedastic disturbances and contemporaneous correlation but with no autocorrelation, the disturbance covariance matrix is assumed to be

$$E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'] = \boldsymbol{\Omega} = \begin{bmatrix} \sigma_{11}\mathbf{I}_{11} & \sigma_{12}\mathbf{I}_{12} & \cdots & \sigma_{1m}\mathbf{I}_{1m} \\ \sigma_{21}\mathbf{I}_{21} & \sigma_{22}\mathbf{I}_{22} & \cdots & \sigma_{2m}\mathbf{I}_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1}\mathbf{I}_{m1} & \sigma_{m2}\mathbf{I}_{m2} & \cdots & \sigma_{mm}\mathbf{I}_{mm} \end{bmatrix}$$

where σ_{ii} is the variance of the disturbances for panel i , σ_{ij} is the covariance of the disturbances between panel i and panel j when the panels' periods are matched, and \mathbf{I} is a T_i by T_i identity matrix with balanced panels. The panels need not be balanced for `xtpcse`, but the expression for the covariance of the disturbances will be more general if they are unbalanced.

This could also be written as

$$E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'] = \boldsymbol{\Sigma}_{m \times m} \otimes \mathbf{I}_{T_i \times T_i}$$

where $\boldsymbol{\Sigma}$ is the panel-by-panel covariance matrix and \mathbf{I} is an identity matrix.

See [XT] [xtgls](#) for a full taxonomy and description of possible disturbance covariance structures.

`xtpcse` and `xtgls` follow two different estimation schemes for this family of models. `xtpcse` produces OLS estimates of the parameters when no autocorrelation is specified, or Prais–Winsten (see [TS] [prais](#)) estimates when autocorrelation is specified. If autocorrelation is specified, the estimates of the parameters are conditional on the estimates of the autocorrelation parameter(s). The estimate of the variance–covariance matrix of the parameters is asymptotically efficient under the assumed covariance structure of the disturbances and uses the FGLS estimate of the disturbance covariance matrix; see [Kmenta \(1997, 121\)](#).

`xtgls` produces full FGLS parameter and variance–covariance estimates. These estimates are conditional on the estimates of the disturbance covariance matrix and are conditional on any autocorrelation parameters that are estimated; see [Kmenta \(1997\)](#), [Greene \(2018\)](#), [Davidson and MacKinnon \(1993\)](#), or [Judge et al. \(1985\)](#).

Both estimators are consistent, as long as the conditional mean ($\mathbf{x}_{it}\beta$) is correctly specified. If the assumed covariance structure is correct, FGLS estimates produced by `xtgls` are more efficient. [Beck and Katz \(1995\)](#) have shown, however, that the full FGLS variance–covariance estimates are typically unacceptably optimistic (anticonservative) when used with the type of data analyzed by most social scientists—10–20 panels with 10–40 periods per panel. They show that the OLS or Prais–Winsten estimates with PCSEs have coverage probabilities that are closer to nominal.

Because the covariance matrix elements, σ_{ij} , are estimated from panels i and j , using those observations that have common time periods, estimators for this model achieve their asymptotic behavior as the T_i s approach infinity. In contrast, the random- and fixed-effects estimators assume a different model and are asymptotic in the number of panels m ; see [XT] [xtreg](#) for details of the random- and fixed-effects estimators.

Although `xtpcse` allows other disturbance covariance structures, the term PCSE, as used in the literature, refers specifically to models that are both heteroskedastic and contemporaneously correlated across panels, with or without autocorrelation.

► Example 1: Controlling for heteroskedasticity and cross-panel correlation

[Grunfeld and Griliches \(1960\)](#) analyzed a company’s current-year gross investment (`invest`) as determined by the company’s prior year market value (`mvalue`) and the prior year’s value of the company’s plant and equipment (`kstock`). The dataset includes 10 companies over 20 years, from 1935 through 1954, and is a classic dataset for demonstrating cross-sectional time-series analysis. [Greene \(2012, 1112\)](#) reproduces the dataset.

To use `xtpcse`, the data must be organized in “long form”; that is, each observation must represent a record for a specific company at a specific time; see [D] [reshape](#). In the Grunfeld data, `company` is a categorical variable identifying the company, and `year` is a variable recording the year. Here are the first few records:

```
. use https://www.stata-press.com/data/r18/grunfeld
. list in 1/5
```

| | company | year | invest | mvalue | kstock | time |
|----|---------|------|--------|--------|--------|------|
| 1. | 1 | 1935 | 317.6 | 3078.5 | 2.8 | 1 |
| 2. | 1 | 1936 | 391.8 | 4661.7 | 52.6 | 2 |
| 3. | 1 | 1937 | 410.6 | 5387.1 | 156.9 | 3 |
| 4. | 1 | 1938 | 257.7 | 2792.2 | 209.2 | 4 |
| 5. | 1 | 1939 | 330.8 | 4313.2 | 203.4 | 5 |

To compute PCSEs, Stata must be able to identify the panel to which each observation belongs and be able to match the periods across the panels. We tell Stata how to do this matching by specifying

the panel and time variables with `xtset`; see [XT] `xtset`. Because the data are annual, we specify the `yearly` option.

```
. xtset company year, yearly
Panel variable: company (strongly balanced)
Time variable: year, 1935 to 1954
Delta: 1 year
```

We can obtain OLS parameter estimates for a linear model of `invest` on `mvalue` and `kstock` while allowing the standard errors (and variance–covariance matrix of the estimates) to be consistent when the disturbances from each observation are not independent. Specifically, we want the standard errors to be robust to each company having a different variance of the disturbances and to each company’s observations being correlated with those of the other companies through time.

This model is fit in Stata by typing

```
. xtpcse invest mvalue kstock
Linear regression, correlated panels corrected standard errors (PCSEs)
Group variable:   company           Number of obs   =       200
Time variable:   year              Number of groups =       10
Panels:          correlated (balanced)  Obs per group:
Autocorrelation: no autocorrelation           min =       20
                                                avg =       20
                                                max =       20
Estimated covariances =       55           R-squared       =   0.8124
Estimated autocorrelations =       0       Wald chi2(2)    =   637.41
Estimated coefficients =       3           Prob > chi2     =   0.0000
```

| invest | Panel-corrected | | | | |
|--------|-----------------|-----------|-------|-------|----------------------|
| | Coefficient | std. err. | z | P> z | [95% conf. interval] |
| mvalue | .1155622 | .0072124 | 16.02 | 0.000 | .101426 .1296983 |
| kstock | .2306785 | .0278862 | 8.27 | 0.000 | .1760225 .2853345 |
| _cons | -42.71437 | 6.780965 | -6.30 | 0.000 | -56.00482 -29.42392 |

◀

► Example 2: Comparing the FGLS and PCSE approaches

`xtgls` will produce more efficient FGLS estimates of the models’ parameters, but with the disadvantage that the standard error estimates are conditional on the estimated disturbance covariance. [Beck and Katz \(1995\)](#) argue that the improvement in power using FGLS with such data is small and that the standard error estimates from FGLS are unacceptably optimistic (anticonservative).

The FGLS model is fit by typing

```
. xtglm invest mvalue kstock, panels(correlated)
Cross-sectional time-series FGLS regression
Coefficients: generalized least squares
Panels: heteroskedastic with cross-sectional correlation
Correlation: no autocorrelation
Estimated covariances = 55 Number of obs = 200
Estimated autocorrelations = 0 Number of groups = 10
Estimated coefficients = 3 Time periods = 20
Wald chi2(2) = 3738.07
Prob > chi2 = 0.0000
```

| invest | Coefficient | Std. err. | z | P> z | [95% conf. interval] | |
|--------|-------------|-----------|--------|-------|----------------------|-----------|
| mvalue | .1127515 | .0022364 | 50.42 | 0.000 | .1083683 | .1171347 |
| kstock | .2231176 | .0057363 | 38.90 | 0.000 | .2118746 | .2343605 |
| _cons | -39.84382 | 1.717563 | -23.20 | 0.000 | -43.21018 | -36.47746 |

The coefficients between the two models are close; the constants differ substantially, but we are generally not interested in the constant. As Beck and Katz observed, the standard errors for the FGLS model are 50%–100% smaller than those for the OLS model with PCSE.

If we were also concerned about autocorrelation of the disturbances, we could obtain a model with a common AR(1) parameter by specifying correlation(ar1).

```
. xtpcse invest mvalue kstock, correlation(ar1)
note: estimates of rho outside [-1,1] bounded to be in the range [-1,1].
Prais-Winsten regression, correlated panels corrected standard errors (PCSEs)
Group variable: company Number of obs = 200
Time variable: year Number of groups = 10
Panels: correlated (balanced) Obs per group:
Autocorrelation: common AR(1) min = 20
avg = 20
max = 20
Estimated covariances = 55 R-squared = 0.5468
Estimated autocorrelations = 1 Wald chi2(2) = 93.71
Estimated coefficients = 3 Prob > chi2 = 0.0000
```

| invest | Panel-corrected | | z | P> z | [95% conf. interval] | |
|--------|-----------------|-----------|-------|-------|----------------------|----------|
| | Coefficient | std. err. | | | | |
| mvalue | .0950157 | .0129934 | 7.31 | 0.000 | .0695492 | .1204822 |
| kstock | .306005 | .0603718 | 5.07 | 0.000 | .1876784 | .4243317 |
| _cons | -39.12569 | 30.50355 | -1.28 | 0.200 | -98.91154 | 20.66016 |
| rho | .9059774 | | | | | |

The estimate of the autocorrelation parameter is high (0.906), and the standard errors are larger than for the model without autocorrelation, which is to be expected if there is autocorrelation.

▶ Example 3: Controlling for cross-panel correlation and autocorrelation

Let's estimate panel-specific autocorrelation parameters and change the method of estimating the autocorrelation parameter to the one typically used to estimate autocorrelation in time-series analysis.

```
. xtpcse invest mvalue kstock, correlation(psar1) rhotype(tscorr)
Prais-Winsten regression, correlated panels corrected standard errors (PCSEs)
Group variable:      company          Number of obs   =       200
Time variable:      year              Number of groups =       10
Panels:              correlated (balanced)  Obs per group:
Autocorrelation:    panel-specific AR(1)    min =           20
                                                           avg =           20
                                                           max =           20
Estimated covariances =           55      R-squared        =       0.8670
Estimated autocorrelations =           10      Wald chi2(2)     =       444.53
Estimated coefficients =           3          Prob > chi2      =       0.0000
```

| invest | Panel-corrected | | | | | [95% conf. interval] |
|--------|-----------------|-----------|----------|--------|--------------|----------------------|
| | Coefficient | std. err. | z | P> z | | |
| mvalue | .1052613 | .0086018 | 12.24 | 0.000 | .0884021 | .1221205 |
| kstock | .3386743 | .0367568 | 9.21 | 0.000 | .2666322 | .4107163 |
| _cons | -58.18714 | 12.63687 | -4.60 | 0.000 | -82.95496 | -33.41933 |
| rhos = | .5135627 | .87017 | .9023497 | .63368 | .8571502 ... | .8752707 |

Beck and Katz (1995, 121) make a case against estimating panel-specific AR parameters, as opposed to one AR parameter for all panels.

◀

▶ Example 4: Controlling for heteroskedasticity only; not quite PCSEs

We can also diverge from PCSEs to estimate standard errors that are panel corrected, but only for panel-level heteroskedasticity; that is, each company has a different variance of the disturbances. Allowing also for autocorrelation, we would type

```
. xtpcse invest mvalue kstock, correlation(ar1) hetonly
note: estimates of rho outside [-1,1] bounded to be in the range [-1,1].
Prais-Winsten regression, heteroskedastic panels corrected standard errors
Group variable:      company          Number of obs   =       200
Time variable:      year              Number of groups =       10
Panels:              heteroskedastic (balanced)  Obs per group:
Autocorrelation:    common AR(1)          min =           20
                                                           avg =           20
                                                           max =           20
Estimated covariances =           10      R-squared        =       0.5468
Estimated autocorrelations =           1      Wald chi2(2)     =       91.72
Estimated coefficients =           3          Prob > chi2      =       0.0000
```

| invest | Het-corrected | | | | | [95% conf. interval] |
|--------|---------------|-----------|-------|-------|-----------|----------------------|
| | Coefficient | std. err. | z | P> z | | |
| mvalue | .0950157 | .0130872 | 7.26 | 0.000 | .0693653 | .1206661 |
| kstock | .306005 | .061432 | 4.98 | 0.000 | .1856006 | .4264095 |
| _cons | -39.12569 | 26.16935 | -1.50 | 0.135 | -90.41666 | 12.16529 |
| rho | .9059774 | | | | | |

With this specification, we do not obtain what are referred to in the literature as PCSEs. These standard errors are in the same spirit as PCSEs but are from the asymptotic covariance estimates of OLS without allowing for contemporaneous correlation.



Stored results

xtpcse stores the following in `e()`:

Scalars

| | |
|-------------------------|--|
| <code>e(N)</code> | number of observations |
| <code>e(N_g)</code> | number of groups |
| <code>e(N_gaps)</code> | number of gaps |
| <code>e(n_cf)</code> | number of estimated coefficients |
| <code>e(n_cv)</code> | number of estimated covariances |
| <code>e(n_cr)</code> | number of estimated correlations |
| <code>e(n_sigma)</code> | observations used to estimate elements of Σ |
| <code>e(mss)</code> | model sum of squares |
| <code>e(df)</code> | degrees of freedom |
| <code>e(df_m)</code> | model degrees of freedom |
| <code>e(rss)</code> | residual sum of squares |
| <code>e(g_min)</code> | smallest group size |
| <code>e(g_avg)</code> | average group size |
| <code>e(g_max)</code> | largest group size |
| <code>e(r2)</code> | R^2 |
| <code>e(chi2)</code> | χ^2 |
| <code>e(p)</code> | p -value for model test |
| <code>e(rmse)</code> | root mean squared error |
| <code>e(rank)</code> | rank of $e(V)$ |
| <code>e(rc)</code> | return code |

Macros

| | |
|------------------------------|--------------------------------------|
| <code>e(cmd)</code> | xtpcse |
| <code>e(cmdline)</code> | command as typed |
| <code>e(depvar)</code> | name of dependent variable |
| <code>e(ivar)</code> | variable denoting groups |
| <code>e(tvar)</code> | variable denoting time within groups |
| <code>e(wtype)</code> | weight type |
| <code>e(wexp)</code> | weight expression |
| <code>e(title)</code> | title in estimation output |
| <code>e(panels)</code> | contemporaneous covariance structure |
| <code>e(corr)</code> | correlation structure |
| <code>e(rhotype)</code> | type of estimated correlation |
| <code>e(rho)</code> | ρ |
| <code>e(cons)</code> | noconstant or "" |
| <code>e(missmeth)</code> | casewise or pairwise |
| <code>e(balance)</code> | balanced or unbalanced |
| <code>e(chi2type)</code> | Wald; type of model χ^2 test |
| <code>e(vctype)</code> | title used to label Std. err. |
| <code>e(properties)</code> | b V |
| <code>e(predict)</code> | program used to implement predict |
| <code>e(marginsok)</code> | predictions allowed by margins |
| <code>e(marginsnotok)</code> | predictions disallowed by margins |
| <code>e(asbalanced)</code> | factor variables fvset as asbalanced |
| <code>e(asobserved)</code> | factor variables fvset as asobserved |

Matrices

| | |
|------------------------|---|
| <code>e(b)</code> | coefficient vector |
| <code>e(Sigma)</code> | $\widehat{\Sigma}$ matrix |
| <code>e(rhomat)</code> | vector of autocorrelation parameter estimates |
| <code>e(V)</code> | variance-covariance matrix of the estimators |

Functions

| | |
|------------------------|-------------------------|
| <code>e(sample)</code> | marks estimation sample |
|------------------------|-------------------------|

Methods and formulas

If no autocorrelation is specified, the parameters β are estimated by OLS; see [R] `regress`. If autocorrelation is specified, the parameters β are estimated by Prais–Winsten; see [TS] `prais`.

When autocorrelation with panel-specific coefficients of correlation is specified (by using option `correlation(psar1)`), each panel-level ρ_i is computed from the residuals of an OLS regression across all panels; see [TS] `prais`. When autocorrelation with a common coefficient of correlation is specified (by using option `correlation(ar1)`), the common correlation coefficient is computed as

$$\rho = \frac{\rho_1 + \rho_2 + \cdots + \rho_m}{m}$$

where ρ_i is the estimated autocorrelation coefficient for panel i and m is the number of panels.

The covariance of the OLS or Prais–Winsten coefficients is

$$\text{Var}(\beta) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$$

where $\mathbf{\Omega}$ is the full covariance matrix of the disturbances.

When the panels are balanced, we can write $\mathbf{\Omega}$ as

$$\mathbf{\Omega} = \mathbf{\Sigma}_{m \times m} \otimes \mathbf{I}_{T_i \times T_i}$$

where $\mathbf{\Sigma}$ is the m by m panel-by-panel covariance matrix of the disturbances; see [Remarks and examples](#).

`xtpcse` estimates the elements of $\mathbf{\Sigma}$ as

$$\widehat{\Sigma}_{ij} = \frac{\epsilon_i' \epsilon_j}{T_{ij}}$$

where ϵ_i and ϵ_j are the residuals for panels i and j , respectively, that can be matched by period, and where T_{ij} is the number of residuals between the panels i and j that can be matched by time period.

When the panels are balanced (each panel has the same number of observations and all periods are common to all panels), $T_{ij} = T$, where T is the number of observations per panel.

When panels are unbalanced, `xtpcse` by default uses `casewise` selection, in which only those residuals from periods that are common to all panels are used to compute \widehat{S}_{ij} . Here $T_{ij} = T^*$, where T^* is the number of periods common to all panels. When `pairwise` is specified, each \widehat{S}_{ij} is computed using all observations that can be matched by period between the panels i and j .

Acknowledgments

We thank the following people for helpful comments: Nathaniel Beck of the Department of Politics at New York University, Jonathan Katz of the Division of the Humanities and Social Science at California Institute of Technology, and Robert John Franzese Jr. of the Center for Political Studies at the Institute for Social Research at the University of Michigan.

References

- Beck, N. L., and J. N. Katz. 1995. What to do (and not to do) with time-series cross-section data. *American Political Science Review* 89: 634–647. <https://doi.org/10.2307/2082979>.
- Blackwell, J. L., III. 2005. Estimation and testing of fixed-effect panel-data systems. *Stata Journal* 5: 202–207.
- Davidson, R., and J. G. MacKinnon. 1993. *Estimation and Inference in Econometrics*. New York: Oxford University Press.
- Greene, W. H. 2012. *Econometric Analysis*. 7th ed. Upper Saddle River, NJ: Prentice Hall.
- . 2018. *Econometric Analysis*. 8th ed. New York: Pearson.
- Grunfeld, Y., and Z. Griliches. 1960. Is aggregation necessarily bad? *Review of Economics and Statistics* 42: 1–13. <https://doi.org/10.2307/1926089>.
- Hoechle, D. 2007. Robust standard errors for panel regressions with cross-sectional dependence. *Stata Journal* 7: 281–312.
- Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lütkepohl, and T.-C. Lee. 1985. *The Theory and Practice of Econometrics*. 2nd ed. New York: Wiley.
- Kmenta, J. 1997. *Elements of Econometrics*. 2nd ed. Ann Arbor: University of Michigan Press.

Also see

- [XT] **xtpcse postestimation** — Postestimation tools for xtpcse
- [XT] **xtgl** — GLS linear model with heteroskedastic and correlated errors
- [XT] **xtreg** — Fixed-, between-, and random-effects and population-averaged linear models
- [XT] **xtregar** — Fixed- and random-effects linear models with an AR(1) disturbance
- [XT] **xtset** — Declare data to be panel data
- [R] **regress** — Linear regression
- [TS] **newey** — Regression with Newey–West standard errors
- [TS] **prais** — Prais–Winsten and Cochrane–Orcutt regression
- [U] **20 Estimation and postestimation commands**

Stata, Stata Press, and Mata are registered trademarks of StataCorp LLC. Stata and Stata Press are registered trademarks with the World Intellectual Property Organization of the United Nations. Other brand and product names are registered trademarks or trademarks of their respective companies. Copyright © 1985–2023 StataCorp LLC, College Station, TX, USA. All rights reserved.

