**xtpcse — Linear regression with panel-corrected standard errors**

**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] _xtgls_ for the generalized least-squares estimator for these models.

**Quick start**

Linear regression of $y$ on $x_1$ and $x_2$ with panel-corrected standard errors and assuming no within-panel autocorrelation using _xtset_ data

```
xtpcse y x1 x2
```

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]
```

### options Description

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>noconstant</strong></td>
<td>suppress constant term</td>
</tr>
<tr>
<td><strong>correlation(independent)</strong></td>
<td>use independent autocorrelation structure</td>
</tr>
<tr>
<td><strong>correlation(ar1)</strong></td>
<td>use AR1 autocorrelation structure</td>
</tr>
<tr>
<td><strong>correlation(psar1)</strong></td>
<td>use panel-specific AR1 autocorrelation structure</td>
</tr>
<tr>
<td><strong>rho</strong></td>
<td>specify method to compute autocorrelation parameter; seldom used</td>
</tr>
<tr>
<td><strong>rho</strong></td>
<td>weight panel-specific autocorrelations by panel sizes</td>
</tr>
<tr>
<td><strong>hetonly</strong></td>
<td>assume panel-level heteroskedastic errors</td>
</tr>
<tr>
<td><strong>independent</strong></td>
<td>assume independent errors across panels</td>
</tr>
<tr>
<td><strong>by</strong></td>
<td>include only observations with complete cases</td>
</tr>
<tr>
<td><strong>pairwise</strong></td>
<td>include all available observations with nonmissing pairs</td>
</tr>
<tr>
<td><strong>SE</strong></td>
<td>normalize standard errors by $N - k$ instead of $N$</td>
</tr>
<tr>
<td><strong>level(#)</strong></td>
<td>set confidence level; default is <code>level(95)</code></td>
</tr>
<tr>
<td><strong>detail</strong></td>
<td>report list of gaps in time series</td>
</tr>
<tr>
<td><strong>display_options</strong></td>
<td>control columns and column formats, row spacing, line width, display of omitted variables and base and empty cells, and factor-variable labeling</td>
</tr>
<tr>
<td><strong>coeflegend</strong></td>
<td>display legend instead of statistics</td>
</tr>
</tbody>
</table>

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

- **regress** regression using lags; the default
- **freg** regression using leads
- **tscorr** time-series autocorrelation calculation
- **dw** 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 $\rho$ 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.

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

detail specifies that a detailed list of any gaps in the series be reported.
The following option is available with xtpcse but is not shown in the dialog box: coeflegend; see [R] Estimation options.

Remarks and examples


txtcse 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

$$y_{it} = x_{it}\beta + \epsilon_{it}$$

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

This model can also be written panel by panel as

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{bmatrix} \beta + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_m \end{bmatrix}$$

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

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

where $\sigma_{ii}$ is the variance of the disturbances for panel $i$, $\sigma_{ij}$ is the covariance of the disturbances between panel $i$ and panel $j$ when the panels’ periods are matched, and $I$ 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

$$E[\epsilon\epsilon'] = \Sigma_{m \times m} \otimes I_{T_i \times T_i}$$

where $\Sigma$ is the panel-by-panel covariance matrix and $I$ 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 ($x_{it}\beta$) 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, $\sigma_{ij}$, 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/r16/grunfeld
. list in 1/5

<table>
<thead>
<tr>
<th>company</th>
<th>year</th>
<th>invest</th>
<th>mvalue</th>
<th>kstock</th>
<th>time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1935</td>
<td>317.6</td>
<td>3078.5</td>
<td>2.8</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>1936</td>
<td>391.8</td>
<td>4661.7</td>
<td>52.6</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td>1937</td>
<td>410.6</td>
<td>5387.1</td>
<td>156.9</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>1938</td>
<td>257.7</td>
<td>2792.2</td>
<td>209.2</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>1939</td>
<td>330.8</td>
<td>4313.2</td>
<td>203.4</td>
<td>5</td>
</tr>
</tbody>
</table>
```

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 \texttt{xtset}; see \texttt{[XT] xtset}. Because the data are annual, we specify the \texttt{yearly} option.

\begin{verbatim}
. xtset company year, yearly
  \text{panel variable: company (strongly balanced)}
  \text{time variable: year, 1935 to 1954}
  \text{delta: 1 year}
\end{verbatim}

We can obtain OLS parameter estimates for a linear model of \texttt{invest} on \texttt{mvalue} and \texttt{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

\begin{verbatim}
. xtpcse invest mvalue kstock
\end{verbatim}

\textbf{Example 2: Comparing the FGLS and PCSE approaches}

\texttt{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. \textit{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

```
xtgls invest mvalue kstock, panels(correlated)
```

Cross-sectional time-series FGLS regression

Coefficients: generalized least squares
Panels: heteroskedastic with cross-sectional correlation
Correlation: no autocorrelation

<table>
<thead>
<tr>
<th>Estimated covariances</th>
<th>Number of obs</th>
<th>55</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated autocorrelations</td>
<td>Number of groups</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Estimated coefficients</td>
<td>Time periods</td>
<td>3</td>
<td>20</td>
</tr>
</tbody>
</table>

Wald chi2(2) = 3738.07
Prob > chi2 = 0.0000

| invest | Coef. | 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)

| invest | Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|--------|-------|-----------|------|------|----------------------|
| 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)

<table>
<thead>
<tr>
<th></th>
<th>Panel-corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
</tr>
<tr>
<td>mvalue</td>
<td>0.1052613</td>
</tr>
<tr>
<td>kstock</td>
<td>0.3386743</td>
</tr>
<tr>
<td>_cons</td>
<td>-58.18714</td>
</tr>
</tbody>
</table>

Estimation results

```
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
```

Prais-Winsten regression, heteroskedastic panels corrected standard errors

<table>
<thead>
<tr>
<th></th>
<th>Het-corrected</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
</tr>
<tr>
<td>mvalue</td>
<td>0.0950157</td>
</tr>
<tr>
<td>kstock</td>
<td>0.306005</td>
</tr>
<tr>
<td>_cons</td>
<td>-39.12569</td>
</tr>
</tbody>
</table>

Estimation results

```
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**

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

**Macros**

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

**Matrices**

- `e(b)` coefficient vector
- `e(Sigma)` \( \Sigma \) matrix
- `e(rhomat)` vector of autocorrelation parameter estimates
- `e(V)` variance–covariance matrix of the estimators

**Functions**

- `e(sample)` marks estimation sample
Methods and formulas

If no autocorrelation is specified, the parameters $\beta$ are estimated by OLS; see [R] regress. If autocorrelation is specified, the parameters $\beta$ 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 $\rho_i$ 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 $\rho_i$ 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

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

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

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

$$\Omega = \Sigma_{m \times m} \otimes I_{T_i \times T_i}$$

where $\Sigma$ is the $m$ by $m$ panel-by-panel covariance matrix of the disturbances; see Remarks and examples.

xtpcse estimates the elements of $\Sigma$ as

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

where $\epsilon_i$ and $\epsilon_j$ 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 $\hat{\Sigma}_{ij}$. Here $T_{ij} = T^*$, where $T^*$ is the number of periods common to all panels. When pairwise is specified, each $\hat{\Sigma}_{ij}$ 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


Also see

[XT] *xtpcse postestimation* — Postestimation tools for *xtpcse*

[XT] *xtgls* — Fit panel-data models by using GLS

[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