The Stata Journal (2001)
1, Number 1, pp. 101–104

Residual diagnostics for cross-section time series regression models

Christopher F. Baum
Boston College
baum@bc.edu

Abstract. These routines support the diagnosis of groupwise heteroskedasticity and cross-sectional correlation in the context of a regression model fit to pooled cross-section time series (`xt`) data.

Keywords: st0004, fixed effects, groupwise heteroskedasticity, contemporaneous correlation

Syntax

`xttest2`

`xttest3`

These tests are for use with cross-section time series data, following the use of `xtreg, fe` or `xtgls`.

Description

The fixed-effects regression model estimated by `xtreg, fe` invokes the ordinary least squares (OLS) estimator for point and interval estimates under the classical assumptions that the error process is independently and identically distributed. In the pooled cross-section time series context, these assumptions may be violated in several ways.

The error process may be homoskedastic within cross-sectional units, but its variance may differ across units: a condition known as groupwise heteroskedasticity. The `xttest3` command calculates a modified Wald statistic for groupwise heteroskedasticity in the residuals of a fixed-effect regression model, following page 598 of Greene (2000). The null hypothesis specifies that $\sigma_i^2 = \sigma^2$ for $i = 1, \dots, N_g$, where N_g is the number of cross-sectional units. Let $\widehat{\sigma}_i^2 = T_i^{-1} \sum_{t=1}^{T_i} e_{it}^2$ be the estimator of the ith cross-sectional unit's error variance, based upon the T_i residuals e_{it} available for that unit. Then define

$$V_i = T_i^{-1} \left(T_i - 1\right)^{-1} \sum_{t=1}^{T_i} \left(e_{it}^2 - \widehat{\sigma}_i^2\right)^2$$

as the estimated variance of $\widehat{\sigma}_i^2$. The modified Wald test statistic, defined as

$$W = \sum_{i=1}^{N_g} \frac{\left(\widehat{\sigma}_i^2 - \widehat{\sigma}^2\right)^2}{V_i}$$

will be distributed as $\chi^2[N_g]$ under the null hypothesis. Greene's discussion of Lagrange multiplier, likelihood ratio, and standard Wald test statistics points out that these statistics are sensitive to the assumption of normality of the errors. The modified Wald statistic computed here is viable when the assumption of normality is violated, at least in asymptotic terms. In terms of small sample properties, simulations of the test statistic's properties have shown that its power is very low in the context of fixed effects with "large N, small T" panels. In that circumstance, the test should be used with caution.

One generalization of Greene's derivation has been applied to allow for unbalanced panels (in which T_i, the number of observations per cross-sectional unit, is not constant across units). All sums are computed over the actual T_i for the ith cross-sectional unit.

A second deviation from independently and identically distributed errors may arise in the context of contemporaneous correlation of errors across cross-sectional units. These correlations are those exploited by Zellner's seemingly unrelated regression (SUR) estimator, and this test is provided by Stata's `sureg, corr` in that context. `xttest2` tests the hypothesis that the residual correlation matrix, computed over observations common to all cross-sectional units, is an identity matrix of order N_g, where N_g is the number of cross-sectional units. The Lagrange multiplier test statistic is

$$\lambda_{LM} = T \sum_{i=2}^{N_g} \sum_{j=1}^{i-1} r_{ij}^2$$

where r_{ij}^2 is the ijth residual correlation coefficient. The Breusch and Pagan (1980) test statistic is distributed $\chi^2[d]$, where $d = N_g(N_g - 1)/2$, under the null hypothesis of cross-sectional independence.

In the context of an unbalanced panel, the observations used to calculate the correlations entering the test statistic are those available for all cross-sectional units (that is, they are not pairwise correlations). The number of available observations is reported, as is the estimated correlation matrix of the residuals over cross-sectional units.

Since this routine makes use of Stata's matrix language, it cannot compute the test if the number of cross-sectional units in the data exceeds 800 (see help `matsize`). If the current `matsize` is less than the number of cross-sectional units in the data, the same problem will arise, but the user can reset `matsize` as long as the number is less than 800.

Saved results

`xttest2` saves the following scalars in `r()`:

`r(chi2_bp)`	test statistic
`r(n_bp)`	number of complete observations
`r(df_bp)`	degrees of freedom

`xttest3` saves the following scalars in `r()`:

`r(wald)`	test statistic
`r(p)`	p-value
`r(df)`	degrees of freedom

Examples

The Grunfeld investment data (20 years of annual data on five US corporations) are analyzed. Firm 2 is excluded, so that N_g is four.

```
. iis firm

. xtreg i f c if firm !=2, fe

Fixed-effects (within) regression               Number of obs      =        80
Group variable (i) : firm                       Number of groups   =         4

R-sq:  within  = 0.7989                         Obs per group: min =        20
       between = 0.7360                                        avg =      20.0
       overall = 0.7568                                        max =        20

                                                F(2,74)            =    147.01
corr(u_i, Xb)  = -0.0895                        Prob > F           =    0.0000

------------------------------------------------------------------------------
           i |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           f |   .1065863   .0179505     5.94   0.000     .0708193    .1423534
           c |   .3474248    .027327    12.71   0.000     .2929744    .4018751
       _cons |   -72.5259     38.467    -1.89   0.063    -149.1731    4.121279
-------------+----------------------------------------------------------------
     sigma_u |  137.00056
     sigma_e |  77.151807
         rho |  .75922222   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(3, 74) =    62.09               Prob > F = 0.0000

. xttest3

Modified Wald test for groupwise heteroskedasticity
in fixed effect regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (4)  =      279.13
Prob>chi2 =      0.0000
```

```
. xttest2

Correlation matrix of residuals:

         __e1     __e3     __e4     __e5
__e1   1.0000
__e3  -0.0740   1.0000
__e4  -0.2723   0.9032   1.0000
__e5  -0.1825  -0.1330  -0.0967   1.0000

Breusch-Pagan LM test of independence: chi2(6) =    19.115, Pr = 0.0040
Based on 20 complete observations

. xtgls i f c if firm !=2, p(h)
  (output omitted)

. xttest3

Modified Wald test for groupwise heteroskedasticity
in cross-sectional time-series FGLS regression model

H0: sigma(i)^2 = sigma^2 for all i

chi2 (4)  =     5903.67
Prob>chi2 =      0.0000

. xttest2

Correlation matrix of residuals:

         __e1     __e3     __e4     __e5
__e1   1.0000
__e3  -0.1896   1.0000
__e4  -0.3742   0.8963   1.0000
__e5  -0.1052  -0.1417  -0.0973   1.0000

Breusch-Pagan LM test of independence: chi2(6) =    20.400, Pr = 0.0024
Based on 20 complete observations
```

The null hypotheses of each test are decisively rejected. The errors exhibit both groupwise heteroskedasticity and contemporaneous correlation, whether fit by fixed-effects or by feasible GLS estimators.

References

Breusch, T. and A. Pagan. 1980. The LM test and its application to model specification in econometrics. *Review of Economic Studies* 47: 239–254.

Greene, W. 2000. *Econometric Analysis*. Upper Saddle River, NJ: Prentice–Hall.

About the Author

Christopher F. Baum is an associate professor of economics at Boston College, where he co-directs the Minor in Scientific Computation in the College of Arts and Sciences. He is an associate editor of *Computational Economics* and *The Stata Journal*, and serves on the Advisory Board of the Society for Computational Economics. Baum founded and manages the Boston College Statistical Software Components archive at RePEc (*http://repec.org*), the largest Web repository of Stata code.