# Testing for Cross-Sectional Dependence in Panel-Data Models

**Article** *in* The Stata Journal Promoting communications on statistics and Stata · February 2006

DOI: 10.1177/1536867X0600600403 · Source: RePEc



**2 authors:**

Rafael De Hoyos
World Bank



Vasilis Sarafidis
Brunel University London

Draft

# Testing for Cross-sectional Dependence in Panel Data Models

Rafael E. De Hoyos
University of Cambridge

Vasilis Sarafidis
University of Sydney

**Abstract.** This paper describes a new Stata routine, `xtcsd`, for testing for the presence of cross-sectional dependence in panels with a large number of cross-sectional units and a small number of time series observations. The command executes three different testing procedures – namely, Friedman's (1937) test statistic, the statistic proposed by Frees (1995) and the CD test of Pesaran (2004). We illustrate the command by means of an empirical example.

**Keywords:** panel data, cross-sectional dependence

## 1 Introduction

A growing body of the panel data literature comes to the conclusion that panel data sets are likely to exhibit substantial cross-sectional dependence, which may arise due to the presence of common shocks and unobserved components that become part of the error term ultimately, spatial dependence, as well as due to idiosyncratic pair-wise dependence in the disturbances with no particular pattern of common components or spatial dependence. See, for example, Robertson and Symons (2000), Pesaran (2004), Anselin (2001) and Baltagi (2005, section 10.5). One reason for this development may be that during the last few decades we have experienced an ever-increasing economic and financial integration of countries and financial entities, which implies strong interdependencies between cross-sectional units. In microeconomic applications, the propensity of individuals to respond to common 'shocks', or common unobserved factors in a similar manner may be plausibly explained by social norms, neighbourhood effects, herd behaviour and genuinely interdependent preferences.

The impact of cross-sectional dependence in estimation naturally depends on a variety of factors, such as the magnitude of the correlations across cross-sections and the nature of cross-sectional dependence itself. Assuming that cross-sectional dependence is caused by the presence of common factors, which are unobserved (and as a result, the effect of these components is felt through the disturbance term) but they are uncorrelated with the included regressors, the standard fixed-effects (FE) and random effects (RE) estimators are consistent, although not efficient, and the estimated standard errors are biased. In this case, different possibilities arise in estimation. For example, one may chose to rely on standard FE/RE methods and correct the standard errors by following the approach proposed by Driskoll and Kraay (1998). Alternatively, one may attempt to obtain an efficient estimator by using the methods put forward by Robertson and Symons (2000) and Coakley, Fuertes and Smith (2002). On the other hand, if the unobserved components that create interdependencies across cross-sections are correlated with the included regressors, these approaches will not work and the FE and RE estimators will be biased and inconsistent. In this case, one may follow the approach proposed by Pesaran (2006). An alternative method would be to apply an instrumental variables (IV) type approach using standard FE IV, or RE IV estimators. However, in practise, it would be difficult to find instruments that are correlated with the regressors and not correlated with the unobserved factors.

The impact of cross-sectional dependence in dynamic panel estimators is comparatively more severe. In particular, Phillips and Sul (2003) show that if there is sufficient cross-sectional dependence in the data and this is ignored in estimation (as it is commonly done by practitioners), the decrease in estimation efficiency can become so large that, in fact, the pooled least squares estimator may provide little gain over the single equation OLS. This result is important as it implies that if one decides to pool a population of cross-sections that is homogeneous in the slope parameters but ignores cross-sectional dependence, then the efficiency gains that one had hoped to achieve, compared to running individual OLS regressions, may largely diminish.

In a recent paper that deals specifically with short dynamic panel data models, Robertson, Sarafidis and Yamagata (2005) show that if there is cross-sectional dependence in the data the standard GMM procedures designed to correct for Nickell biases are not consistent for T fixed, as N $\rightarrow \infty$. This outcome is striking because the very purpose for using these estimators in dynamic panels is to benefit from their desirable large N-asymptotic properties. In addition, the authors show that cross-sectional dependence may also have an important impact on bias-correction-type procedures that retain the fixed effects model as the underlying procedure but attempt to correct for the bias using either the mean or the median of the distribution of the fixed effects estimator.

The above indicates that testing for cross-sectional dependence is important in estimating panel data models. When the time dimension (T) of the panel is larger than the cross-sectional dimension (N), one may use for these purposes the LM test, developed by Breusch and Pagan (1980), which is readily available in Stata using the command `xttest2`. On the other hand, when $T < N$, the LM test statistic does not enjoy any desirable statistical properties in that it exhibits substantial size distortions.[1] Thus, there is clearly a need for testing for cross-sectional dependence in Stata in cases where N is large and T is small − the most commonly encountered situation in panels.

This paper describes a new Stata command that implements three popular tests for cross-sectional dependence. The tests are valid when $T < N$ and can be used with balanced and unbalanced panel. The remaining of this paper is as follows: the next section describes three statistical procedures designed to test for cross-sectional dependence in large-N small-T panels − namely, Pesaran's (2004) test, Friedman's (1937) statistic and the test statistic proposed by Frees (1995).[2] Section 3 describes the newly developed

1. See Pesaran (2004) or Sarafidis, Yamagata and Robertson (2006).
2. An additional test has been recently advanced by Sarafidis, Yamagata and Robertson (2006), which is relevant in dynamic panel models with exogenous regressors. Since the testing procedure is based

Stata command, `xtcsd`. Section 4 illustrates the use of `xtcsd` by means of an empirical example based on gross product equations using a balanced panel data of states in the US during the period 1970 to 1986. This is a widely referenced data set available from Baltagi's (2005) econometric text book. A final section concludes.

## 2 Tests of Cross-sectional Dependence

Consider the standard panel data model

$$y_{it} = \alpha_i + \beta' \mathbf{x}_{it} + u_{it}, \quad i = 1, ..., N \ \text{ and } \ t = 1, ..., T \tag{1}$$

where $\mathbf{x}_{it}$ is a $K$x1 vector of regressors, $\beta$ is a $K$x1 vector of parameters to be estimated and $\alpha_i$ represent time-invariant individual nuisance parameters. Under the null hypothesis $u_{it}$ is assumed to be independent and identically distributed (i.i.d.) over time-periods and across cross-sectional units. Under the alternative, $u_{it}$ may be correlated across cross-sections but the assumption of no serial-correlation remains.

Thus, the hypothesis of interest is

$$H_0 : \rho_{ij} = \rho_{ji} = cor\left(u_{it}, u_{jt}\right) = 0 \quad \text{for} \ \ i \neq j, \tag{2}$$

vs

$$H_1 : \rho_{ij} = \rho_{ji} \neq 0 \ \text{ for some } \ i \neq j, \tag{3}$$

where $\rho_{ij}$ is the product-moment correlation coefficient of the disturbances and is given by

$$\rho_{ij} = \rho_{ji} = \frac{\sum_{t=1}^{T} u_{it} u_{jt}}{\left(\sum_{t=1}^{T} u_{it}^2\right)^{1/2} \left(\sum_{t=1}^{T} u_{jt}^2\right)^{1/2}} \tag{4}$$

Notice that the number of possible pairings $(u_{it}, u_{jt})$ rises with $N$.

### 2.1 Pesaran's CD test

In the context of seemingly unrelated regressions estimation, Breusch and Pagan (1980) proposed a Lagrange Multiplier (LM) statistic, which is valid for fixed $N$ as $T \to \infty$

on a Sargan's type difference test, which can be obtained in Stata in a straightforward way, this test is not analysed here. For more details, see the reference above.

and is given by

$$LM = T\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\widehat{\rho}_{ij}^{2} \tag{5}$$

where $\widehat{\rho}_{ij}$ is the sample estimate of the pair-wise correlation of the residuals

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t=1}^{T}\hat{u}_{it}\hat{u}_{jt}}{\left(\sum_{t=1}^{T}\hat{u}_{it}^{2}\right)^{1/2}\left(\sum_{t=1}^{T}\hat{u}_{jt}^{2}\right)^{1/2}} \tag{6}$$

and $\hat{u}_{it}$ is the estimate of $u_{it}$ in (1). $LM$ is asymptotically distributed as chi-squared with $N(N-1)/2$ degrees of freedom under the null hypothesis of interest. However, this test is likely to exhibit substantial size distortions in cases where $N$ is large and $T$ is finite − a situation that is commonly encountered in empirical applications, primarily due to the fact that the LM statistic is not correctly centered for finite $T$ and the bias is likely to get worse with $N$ large.

Pesaran (2004) has proposed the following alternative:

$$CD = \sqrt{\frac{2T}{N(N-1)}}\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\hat{\rho}_{ij}\right) \tag{7}$$

and showed that under the null hypothesis of no cross-sectional dependence $CD \xrightarrow{d} N(0,1)$ for $N \to \infty$ and $T$ sufficiently large.

Unlike the LM statistic, the CD statistic has exactly mean at zero for fixed values of $T$ and $N$, under a wide range of panel data models, including heterogeneous models, non-stationary models and dynamic panels.

In the case of unbalanced panels, Pesaran (2004) proposes a slightly modified version of equation 7, which is given by

$$CD = \sqrt{\frac{2}{N(N-1)}}\left(\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}\sqrt{T_{ij}}\hat{\rho}_{ij}\right) \tag{8}$$

where $T_{ij} = \#(T_i \cap T_j)$ (i.e. the number of common time series observations between units $i$ and $j$),

$$\hat{\rho}_{ij} = \hat{\rho}_{ji} = \frac{\sum_{t\in T_i\cap T_j}\left(\hat{u}_{it}-\overline{\hat{u}}_{i}\right)\left(\hat{u}_{jt}-\overline{\hat{u}}_{j}\right)}{\left[\sum_{t\in T_i\cap T_j}\left(\hat{u}_{it}-\overline{\hat{u}}_{i}\right)^{2}\right]^{1/2}\left[\sum_{t\in T_i\cap T_j}\left(\hat{u}_{jt}-\overline{\hat{u}}_{j}\right)^{2}\right]^{1/2}} \tag{9}$$

and

$$\overline{\hat{u}}_i = \frac{\sum_{t \in T_i \cap T_j} \hat{u}_{it}}{\# \left(T_i \cap T_j\right)} \tag{10}$$

The modified statistic accounts for the fact that the residuals for subsets of $t$ are not necessarily mean zero.

## 2.2 Friedman's test

Friedman (1937) proposed a non-parametric test based on Spearman's rank correlation coefficient, which can be thought of as the regular product-moment correlation coefficient, that is, in terms of proportion of variability accounted for, except that Spearman's rank correlation coefficient is computed from ranks. In particular, defining $\{r_{i,1}, ..., r_{i,T}\}$ to be the ranks of $\{u_{i,1}, ..., u_{i,T}\}$ (such that the average rank is $(T + 1/2)$), Spearman's rank correlation coefficient equals

$$r_{ij} = r_{ji} = \frac{\sum_{t=1}^{T} \left(r_{i,t} - (T + 1/2)\right) \left(r_{j,t} - (T + 1/2)\right)}{\sum_{t=1}^{T} \left(r_{i,t} - (T + 1/2)\right)^2} \tag{11}$$

Friedman' statistic is based on the average Spearman's correlation and is given by

$$R_{AVE} = \frac{2}{N\left(N-1\right)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \widehat{r}_{ij} \tag{12}$$

where $\widehat{r}_{ij}$ is the sample estimate of the rank correlation coefficient of the residuals. Large values of $R_{AVE}$ indicate the presence of non-zero cross-sectional correlations. Friedman showed that $FR = \left[(T-1)\left((N-1)\, R_{AVE} + 1\right)\right]$ is asymptotically chi-squared distributed with $T-1$ degrees of freedom, for fixed $T$ as $N$ gets large. Notice that originally Friedman devised the test statistic $FR$ in order to determine the equality of treatment in a two-way analysis of variance.

Both the $CD$ and $R_{AVE}$ share a common weakness in that they both involve the sum of the pair-wise correlation coefficients of the residual matrix, rather than the sum of the squared correlations used in the LM test. This implies that these tests are likely to miss out cases of cross-sectional dependence where the sign of the correlations is alternating $-$ that is, where there are large positive and negative correlations in the residuals, which cancel each other out when averaging. Consider, for example, the following error structure of $u_{it}$ under $H_1$ :

$$u_{it} = \phi_i f_t + \varepsilon_{it} \tag{13}$$

where $f_t$ represents the unobserved factor that generates cross-sectional dependence, $\phi_i$

indicates the impact of the factor on unit $i$ and $\varepsilon_{it}$ is a pure idiosyncratic error with $f_t \sim i.i.d\left(0, \sigma_f^2\right)$, $\phi_i \sim i.i.d\left(0, \sigma_\phi^2\right)$ and $\varepsilon_{it} \sim i.i.d\left(0, \sigma_\varepsilon^2\right)$. In this case, we have

$$cor\left(u_{it}, u_{jt}\right) = \frac{cov\left(u_{it}, u_{jt}\right)}{\sqrt{var\left(u_{it}\right)}\sqrt{var\left(u_{jt}\right)}} = \frac{E\left(u_{it}\right)\left(u_{jt}\right)}{\sqrt{E\left[u_{it}\right]^2}\sqrt{E\left[u_{jt}\right]^2}} = 0 \tag{14}$$

and thereby the $CD$ and $R_{AVE}$ statistics converge to 0 even if $f_t \neq 0$ and $\phi_i \neq 0$ for some $i$. This implies that under alternative hypotheses of cross-sectional dependence in the disturbances with large positive and negative correlations but with $E\left(\phi_i\right) = 0$, these tests would lack power and as a result they may not be reliable.

## 2.3 Frees' test

Frees (1995, 2004) proposed a statistic that is not subject to this drawback.[3] In particular, the statistic is based on the sum of the squared rank correlation coefficients and equals

$$R_{AVE}^2 = \frac{2}{N\left(N-1\right)} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \widehat{r}_{ij}^2 \tag{15}$$

As shown by Frees, a function of this statistic follows a joint distribution of two independently drawn $\chi^2$ variables. In particular, Frees shows that

$$\begin{aligned} FRE &= N\left(R_{AVE}^2 - \left(T-1\right)^{-1}\right) \xrightarrow{d} Q = a\left(T\right)\left(x_{1,T-1}^2 - \left(T-1\right)\right) \\ &\quad + b\left(T\right)\left(x_{2,T(T-3)/2}^2 - T\left(T-3\right)/2\right) \end{aligned} \tag{16}$$

where $x_{1,T-1}^2$ and $x_{2,T(T-3)/2}^2$ are independently $\chi^2$ random variables with $T-1$ and $T\left(T-3\right)/2$ degrees of freedom respectively, $a\left(T\right) = 4\left(T+2\right)/\left(5\left(T-1\right)^2\left(T+1\right)\right)$ and $b\left(T\right) = 2\left(5T+6\right)/\left(5T\left(T-1\right)\left(T+1\right)\right)$. Thus, the null hypothesis is rejected if $R_{AVE}^2 > \left(T-1\right)^{-1} + Q_q/N$, where $Q_q$ is the appropriate quantile of the $Q$ distribution.

The $Q$ distribution is a (weighted) sum of two chi-squared distributed random variables and depends on the size of T. Hence, computation of the appropriate quantiles may be quite tedious. In cases where $T$ is not small, Frees suggests using the normal approximation to the $Q$ distribution by computing the variance of $Q$. In other words,

3. The testing procedure proposed by Sarafidis, Yamagata and Robertson (2006) is not subject to this drawback either.

we can make use of the following result:

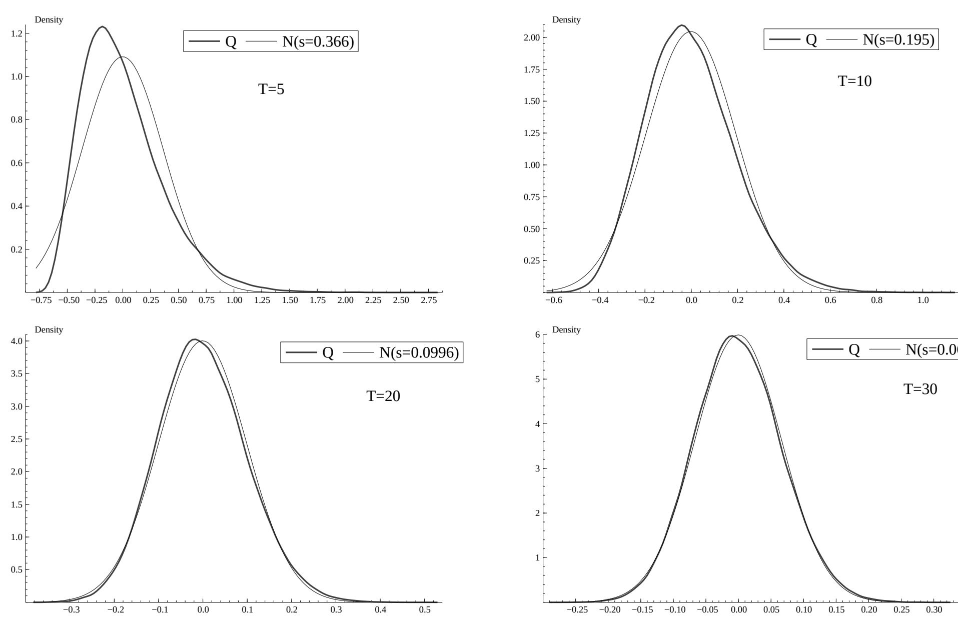
$$\frac{FRE}{\sqrt{Var\left(Q\right)}} \overset{approximately}{\sim} N\left(0,1\right) \tag{17}$$

where

$$Var\left(Q\right) = \frac{32}{25}\frac{\left(T+2\right)^2}{\left(T-1\right)^3\left(T+1\right)^2} + \frac{4}{5}\frac{\left(5T+6\right)^2\left(T-3\right)}{T\left(T-1\right)^2\left(T+1\right)^2} \tag{18}$$

The accuracy of the normal approximation is illustrated in the following diagram, which illustrates the density of Q for different values of $T$:

Figure 1: The normal approximation to the Q distribution (s denotes the standard deviation).




As we can see, for small values of T the normal approximation to the Q distribution is poor. However, for T as large as 30, the approximation does well. Notice that contrary to Pesaran's CD test, the tests by Frees and Friedman have been originally devised for static panels and the finite sample properties of the tests in dynamic panels have not been investigated yet.

# 3 The xtcsd Command

The new Stata command, `xtcsd`, tests for the presence of cross-sectional dependence in fixed effects and random effects panel data models. The command is suitable for cases where $T$ is small as $N \to \infty$. It therefore complements the existing Breusch-Pagan LM test written by Christopher Baum, `xttest2` which is valid for small $N$ as $T \to \infty$. By making available a series of tests for cross-sectional dependence for cases where $N$ is large and $T$ is small, `xtcsd` closes an important gap in applied research.

## 3.1 Syntax

The syntax of `xtcsd` is the following:

`xtcsd` [ , `pesaran` `friedman` `frees` `abs` `show` ]

As it is the case with all other Stata cross-sectional time-series (`xt`) commands, the data needs to be `tsset` before using `xtcsd`. `xtcsd` is a post-estimation command valid for use after running either a Fixed-effects or a Random-effects model.

## 3.2 Options

`pesaran` performs the CD test developed by Pesaran (2004) as explained in section 2.1. In the context of balanced panels, option `pesaran` estimates equation 7. In the case of unbalanced panels, `pesaran` estimates equation 8. The CD statistic is normally distributed under the null hypothesis (equation 2) for $T_i > k+1$, and $T_{ij} > 2$ and sufficiently large N. Therefore there must be enough cross-sectional units with common points in time to be able to implement the test.

`friedman` performs Friedman's test for cross-sectional dependence using the non-parametric chi-square distributed $R_{AVE}$ statistic (see section 2.2). For unbalanced panels Friedman's test uses only the observations available for all cross-sectional units.

`frees` test for cross-sectional dependence using Frees' Q distribution (T-asymptotically distributed). For unbalanced panels Frees' test uses only the observations available for all cross-sectional units. For $T > 30$ `frees` uses a normal approximation to obtain the critical values of the $Q$ distribution.

`abs` computes the average absolute value of the off-diagonal elements of the cross-sectional correlation matrix of residuals. This is useful to identify cases of cross-sectional dependence where the sign of the correlations is alternating with the likely result of making the `pesaran` and `friedman` tests unreliable (see section 2.2).

`show` shows the cross-sectional correlation matrix of residuals.

## 4 An Application

We illustrate the use of `xtcsd` by means of an empirical example, which is taken from Baltagi (2001, page 25). The example refers to a Cobb-Douglas production function relationship investigating the productivity of public capital in private production. The data set consists of a balanced panel of 48 US states, each observed over a period of 17 years (1970 to 1986). This data set and also some explanatory notes can be found on the Wiley web site.[4]

Following Munnell (1990) and Baltagi and Pinnoi (1995), Baltagi (2001) considers the following relationship:

$$\ln gsp_{it} = \alpha + \beta_1 \ln pcap_{it} + \beta_2 \ln pc_{it} + \beta_3 \ln emp_{it} + \beta_4 unemp_{it} + u_{it} \tag{19}$$

where $gsp_{it}$ denotes gross product in state $i$ at time $t$; *pcap* denotes public capital including highways and streets, water and sewer facilities and other public buildings; *pc* denotes the stock of private capital; *emp* is labor input measured as employment in non-agricultural payrolls; and *unemp* is the state unemployment rate included to capture business cycle effects.

We begin the exercise by downloading the data and declaring that it has a panel data format:

```
. use "http://www.econ.cam.ac.uk/phd/red29/xtcsd_baltagi.dta"
. tsset id t
       panel variable:  id, 1 to 48
        time variable:  t, 1970 to 1986
```

Once the data set is ready for undertaking panel data analysis, we run a version of equation 19 where we assume that $u_{it}$ is formed by a combination of a fixed component inherent to the state and a random component that captures pure noise. The results of the model using the fixed effects estimator, also reported in page 25 of Baltagi (2001), are given below:

```
. xtreg  lngsp lnpcap lnpc lnemp unemp, fe
Fixed-effects (within) regression               Number of obs      =       816
Group variable (i): id                          Number of groups   =        48
R-sq:  within  = 0.9413                         Obs per group: min =        17
       between = 0.9921                                        avg =      17.0
       overall = 0.9910                                        max =        17
                                                F(4,764)           =   3064.81
corr(u_i, Xb)  = 0.0608                         Prob > F           =    0.0000
------------------------------------------------------------------------------
       lngsp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
```

4. The database in plain format is available from http://www.wiley.com/legacy/wileychi/baltagi/supp/PRODUC.prn; type `net from http://www.econ.cam.ac.uk/phd/red29/` in the Stata command browser to get the data in Stata format.

```
      lnpcap |  -.0261493   .0290016    -0.90   0.368    -.0830815    .0307829
        lnpc |   .2920067   .0251197    11.62   0.000     .2426949    .3413185
       lnemp |   .7681595   .0300917    25.53   0.000     .7090872    .8272318
       unemp |  -.0052977   .0009887    -5.36   0.000    -.0072387   -.0033568
       _cons |   2.352898   .1748131    13.46   0.000     2.009727    2.696069
-------------+----------------------------------------------------------------
     sigma_u |  .09057293
     sigma_e |  .03813705
         rho |   .8494045   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(47, 764) =    75.82            Prob > F = 0.0000
```

According to the results, once we account for State fixed effects, public capital has no effect upon state gross product in the US. An assumption implicit in estimating equation 19 is that the cross-sectional units are independent. The `xtcsd` command allows us to test the following hypothesis:

$$Ho: \textit{Cross-sectional Independence}$$

To test the this hypothesis, we use the `xtcsd` command after estimating the above panel data model. We initially employ Pesaran's (2004) CD test:

```
. xtcsd, pesaran abs


Pesaran's test of cross sectional independence =    30.368, Pr = 0.0000

Average absolute value of the off-diagonal elements =     0.442
```

As we can see, the CD test strongly rejects the null hypothesis of no cross-sectional dependence at least at the 1% level of significance. Although it is not the case here, a possible drawback of the CD test is that by adding up positive and negative correlations it might undermine the cross-sectional dependence present in the data. Including the `abs` option in the `xtcsd` command we can get the average *absolute* correlation between the cross-sectional units. In our case the average absolute correlation is 0.439, which is a very high value. Hence there is enough evidence suggesting the presence of cross-sectional dependence in model 19 under a fixed effects assumption.

Next we corroborate these results using the remaining two tests explained in section 2, i.e. Frees (1995) and Friedman (1937):

```
. xtcsd, frees


  Frees' test of cross sectional independence =     8.386
|--------------------------------------------------------|
  Critical values from Frees' Q distribution
                      alpha = 0.10 :   0.1521
                      alpha = 0.05 :   0.1996
                      alpha = 0.01 :   0.2928
```

```
.
. xtcsd, friedman


Friedman's test of cross sectional independence =   152.804, Pr = 0.0000
```

As we would have expected from the highly significant results of the CD test, both Frees' and Friedman's tests reject the null of cross-sectional independence. Notice that, since $T \leq 30$, Frees' test provides the critical values for $\alpha = 0.10$, $\alpha = 0.05$ and $\alpha = 0.01$ from the $Q$ distribution. Frees' statistic is beyond the critical value with at least $\alpha = 0.01$.

Baltagi also reports the results of the model using the random effects estimator. The results are shown below:

```
. xtreg  lngsp lnpcap lnpc lnemp unemp,re
Random-effects GLS regression                   Number of obs      =       816
Group variable (i): id                          Number of groups   =        48

R-sq:  within  = 0.9412                         Obs per group: min =        17
       between = 0.9928                                        avg =      17.0
       overall = 0.9917                                        max =        17

Random effects u_i ~ Gaussian                   Wald chi2(4)       =  19131.09
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

------------------------------------------------------------------------------
       lngsp |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lnpcap |   .0044388   .0234173     0.19   0.850    -.0414583    .0503359
        lnpc |   .3105483   .0198047    15.68   0.000     .2717317    .3493649
       lnemp |   .7296705   .0249202    29.28   0.000     .6808278    .7785132
       unemp |  -.0061725   .0009073    -6.80   0.000    -.0079507   -.0043942
       _cons |   2.135411   .1334615    16.00   0.000     1.873831     2.39699
-------------+----------------------------------------------------------------
     sigma_u |   .0826905
     sigma_e |  .03813705
         rho |  .82460109   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

The results of this second model are inline with the previous one, with public capital having no significant effects upon gross state output. We now test for cross-sectional independence using the new random effects specification:

```
. xtcsd, pesaran


Pesaran's test of cross sectional independence =    29.079, Pr = 0.0000

.
. xtcsd, frees


  Frees' test of cross sectional independence =     8.298
|--------------------------------------------------------|
  Critical values from Frees' Q distribution
                    alpha = 0.10 :   0.1521
```

```
                         alpha = 0.05 :   0.1996
                         alpha = 0.01 :   0.2928

.
. xtcsd, friedman


Friedman's test of cross sectional independence =   144.941, Pr = 0.0000
```

The conclusion with respect to the existence or not of cross-sectional dependence in the errors is not altered. The results show that there is enough evidence to reject the null hypothesis of cross-sectional independence. The newly developed `xtcsd` Stata command shows an easy way of performing three popular tests for cross-sectional dependence.

## 5 Conclusion

This paper has described a new Stata post-estimation command, `xtcsd`, which tests for the presence of cross-sectional dependence in fixed and random effects panel data models. The command executes three different testing procedures–namely, Friedman's (1937) test statistic, the statistic proposed by Frees (1995) and the CD test developed by Pesaran (2004). These procedures are valid in cases where $T$ is fixed and $N$ is large. `xtcsd` is capable of performing Pesaran's (2004) CD test for unbalanced panels. The command complements the Stata command `xttest2` which tests for the presence of error cross-sectional when $T$ large and finite $N$. Hence, `xtcsd` closes an important gap in applied research.

## 6 References

[1] Anselin, L. (2001) *"Spatial Econometrics",* Chapter 14 in B. Baltagi , ed., A Companion to Theoretical Econometrics, Blackwell Publishers, Massachuttes.

[2] Arellano, M., and S. Bond (1991) "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations", *Review of Economic Studies* 58, 277-297.

[3] Baltagi, B.H. (2005) "Econometric Analysis of Panel Data", John Wiley and Sons, third edition.

[4] Baltagi, B.H. and N. Pinnoi (1995) "Public Capital Stock and State Productivity Growth: Further Evidence from an Error Components Model", *Empirical Economics,* 20, 351-359.

[5] Breusch, T.S., and A. R. Pagan (1980) "The Lagrange Multiplier test and its application to model specifications in econometrics", *Review of Economic Studies* 47, 239-53.

[6] Frees, E.W. (1995) "Assessing cross-sectional correlation in panel data", *Journal of Econometrics* 69, 393-414.

[7] Frees E.W. (2004) "Longitudinal and Panel Data: Analysis and Applications in the Social Sciences", Cambridge University Press".

[8] Friedman, M. (1937) "The Use of Ranks to Avoid the Assumption of Normality Implicit in the Analysis of Variance", *Journal of the American Statistical Association*, 32, 675-701.

[9] Munnell, A. (1990) "Why Has Productivity Growth Declined? Productivity and Public Investment", *New England Economic Review,* 3-22.

[10] Pesaran, M.H. (2004) "General diagnostic tests for cross section dependence in panels", *Cambridge Working Papers in Economics* No. 0435, Faculty of Economics, University of Cambridge.

[11] Pesaran, M.H. (2006) "Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure", *Econometrica.*

[12] Phillips, P., and D. Sul (2003) "Dynamic panel estimation and homogeneity testing under cross section dependence", *The Econometrics Journal* 6, 217-259.

[13] Robertson, D., and J. Symons (2000) "Factor residuals in SUR regressions: estimating panels allowing for cross sectional correlation", *CEP discussion paper* No. 0473, Centre for Economic Performance, LSE, London.

[14] Robertson, D. Sarafidis, V. and T. Yamagata (2005) "The Impact of Cross-sectional Dependence in Short Dynamic Panel Estimation", *mimeo*, University of Cambridge.

[15] Sarafidis, V., Yamagata, T., and D. Robertson (2006) "A Test of Cross Section Dependence for a Linear Dynamic Panel Model with Regressors", *mimeo*, University of Cambridge.