Robust Cross Sectional Dependence Test in Panel Regression Model

By:

Nurul Sima Md. Shariff, Nor Aishah Hamzah
and
Ahmad Zubaidi Baharumshah

(Paper presented at the ***International Conference on Robust Statistics*** held on 14-19 Jun 2009 in Parma, Italy)

Robust Cross Sectional Dependence Test in Panel Regression Model

Nurul Sima Md.Shariff[a], Nor Aishah Hamzah[a]
Ahmad Zubaidi Baharumshah[b]
[a]Institute of Mathematical Sciences, University of Malaya, 50603 Kuala Lumpur
[b]Faculty of Economic and Administration, University Putra Malaysia, Selangor

Abstract

The central focus in most recent studies on panel data is on the issue of cross sectional dependence; there exist correlations between different groups (cross section) of innovations in the panel. In the presence of cross sectional dependence, we can no longer use the general assumption of independence across units for disturbances. Thus, it is important to test for cross sectional dependence before modeling and estimating panel data takes place in order to avoid the misspecification of the model which subsequently results in invalid tests, bias, inconsistency and inefficiency in parameter estimates.

Difficulty arises when (i) disturbances are correlated across cross section; (ii) data are subjected to outliers/ shocks. It has been reported that the main possible causes are global shocks and unobserved factors that affect the cross sectional dependence (see Cerrato (2001)). However, in reality, it is also possible to have local shocks which affect innovations in the model. The question that usually arises is whether cross dependencies does exist among innovations in the presence of outliers. The usual tests based on Breusch and Pagan (1980) and Pesaran (2004) uses residuals obtained from Ordinary Least Square Fit (OLS) which are subject to influence of outliers. This paper propose robust cross dependency tests of Breusch and Pagan (1980) and Pesaran (2004) to aid cross sectional dependence test in the presence of outliers. Here, robust regression (RREG) model is employed to capture the outliers' effects and filter outlying observation by down weighting the spurious data.

In this paper, we compare the performances of the Breusch and Pagan (1980), Pesaran (2004) tests as well as the proposed methods under varying degree of cross section dependencies via simulation studies. Our results have shown that robust approaches are able to produce reliable test of cross sectional dependency under the conditions considered.

Keywords: Panel Regression, Cross Section Dependence, Local Shocks, Robust Regression, Huber function, Diagnostic Tool.

1. Introduction

For the last few years, many researchers in economics focused on the issues of cross sectional dependence in panel data (see for example: Coakley et. al (2002), Bai and Ng (2002), Philips and Sul (2003), Moon and Perron (2004), Kapetonis and Pesaran (2004), Coakley et al. (2006), Pesaran (2006), Noman (2008), Pesaran et. al (2008) and Sarafidis et. al (2009))[1]. The common problem faced is the presence of cross correlation effects among innovation in panel model. where the assumption of independence across units for disturbances (Beck and Katz (1996) and Cerrato and Sarantis (2007)) are no longer valid; otherwise, the results obtained for the parameter estimation as well as the standard error will be bias, inconsistent and inefficient (Hoyos and Sarafidis (2006)). Some of the researches tend to ignore the problem due to the difficulty involved in estimation and modeling in the presence of cross sectional dependence (Phillips and Sul (2003)). In real situation however, there is a possibility of observing cross dependency when cross sectional units are correlated to each others. For example; if we study the profit achieved by several firms, there will be the unobserved factor influence the firms' profit, which indirectly caused the cross correlations among the firms.

Among the main possible causes for this problem are the global shocks and unobserved factors (Cerrato (2001)). However, in reality, it is possible to have local shocks which affect the innovations in the model. The questions which may arise are; (1) whether cross dependencies does really exist among innovations in the presence of outliers. (2) will the presence of this local shock lead to biasness towards acceptance of the null hypothesis of no cross sectional dependence?

There are many established studies, developed to overcome cross dependence problem. Amongst them are those of Breusch and Pagan (1980), Pesaran (2004), Hoyos and Sarafidis (2006), Frees (1995). The problem with the usual tests such as those of Breusch and Pagan (1980) and Pesaran (2004) is normally due to the use of residuals obtained from Ordinary Squares Fit (OLS) which are known to be influential to outliers.

[1] Some literatures focusing on estimating panel model with the presence of cross sectional dependence and others testing the existence of unit root in the presence of cross dependency among the residuals.

To overcome this problem, the robust regression is often used to detect the influence of outliers.

The paper is organized as follows: Section 2 describes the model and assumptions used in this study. Section 3 discusses the development of a new robust cross dependency tests. Section 4 reports the Monte Carlo simulation studies and in last section, we conclude the paper.

2. Model

Consider the panel model which has the following form:

$$Y_{it} = \alpha_i + \beta_i X_{it} + e_{it} \quad ; i = 1,2,\ldots\ldots,N. \; t = 1,2,\ldots\ldots,T \qquad (1)$$

where α_i, β_i are the unknown parameters and differ across i, X_{it} being the independent variable, e_{it} is the random errors and Y_{it} is the dependent variable.

In the presence of cross dependency, the random errors in (1) will have the following form:

$$e_{it} = \gamma_i f_t + \varepsilon_{it} \quad ; i = 1,2,\ldots\ldots,N. \; t = 1,2,\ldots\ldots,T \qquad (2)$$

where f_t is the t-th latent factor, γ_i are the i-th factor loadings that may be influenced by the factors and ε_{it} is the random errors.

The factor structure is used to represent the error cross section dependence in majority of literatures (Sarafidis et. al (2009)). It may also differ across cross-section depending on the type of cross-section studied.

The assumptions needed prior to testing for cross dependency in panel model are:

i. e_{it} are serially uncorrelated for each i with mean 0 and the variance σ^2, $0 < \sigma_i^2 < \infty$

ii. Under the null hypothesis of no cross dependency error:

$$e_{it} = \sigma_i \varepsilon_{it}$$

where $\varepsilon_{it} \sim IIDN(0,1)$ for all i and t

iii. $E(X_{it}, e_{it}) = 0$, no autocorrelation between the regressors and residuals across time and units; and e_{it} and X_{it} are independent.

iv. $E(e_{it} | X_{i1}, ..., X_{iT}) = 0$, X_{it} are strictly exogeneous.

v. $E(e_{st}, e_{it}) = 0$, where residuals are serially uncorrelated for $s \neq t$.

For a test on no cross dependency in the data, we have the null hypothesis:

H0: $E(e_{it}, e_{jt}) = 0$ for all $i \neq j$; for $i, j = 1,2,.....,N$; $t = 1,2,.....,T$
(no cross dependency)

H1: $E(e_{it}, e_{jt}) \neq 0$ for some $i \neq j$; for $i, j = 1,2,.....,N$; $t = 1,2,.....,T$
(at least two cross section units are dependent)

The standard tests used to detect the presence of cross dependency in panel model are LM test of Bruesch and Pagan (1980) and Pesaran (1980) and these tests are described in next section.

3. Robust Cross Dependency Test

3.1 RREG with Huber function

Here, we apply the filter function of Huber via RREG, to the fitted residuals in order to eliminate the effects of outliers. The effects of outliers will be down weighted by through the Huber filter function. Both formula for CD_{lm} and PCD are modified as:

i. Robust LM Test 1

$$RCD1 = T\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \hat{\rho}^2_{\psi(ij)} \quad (3)$$

ii. Robust PCD Test 1

$$RPCD1 = \sqrt{\frac{2T}{N(N-1)}}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \hat{\rho}_{\psi(ij)} \quad (4)$$

where $$\hat{\rho}_{\psi(ij)} = \hat{\rho}_{\psi(ji)} = \frac{\sum_{t=1}^{T}\psi(\hat{u}_{it})\psi(\hat{u}_{jt})}{\left(\sum_{t=1}^{T}\psi(\hat{u}_{it})^2\right)^{1/2}\left(\sum_{t=1}^{T}\psi(\hat{u}_{jt})^2\right)^{1/2}} \quad (5)$$

and $\hat{u}_{it}$ is standardized residuals given by

$$\hat{u}_{it} = \frac{\hat{e}'_{it}}{k\,\underset{i}{med}\left|\hat{e}'_{it} - \underset{j}{med}(\hat{e}'_{jt})\right|} \quad (8)$$

The $\hat{e}'_{it}$ is the fitted residual obtained from RREG while k is a constant scale factor. Under the normal distribution, k is set to be 1.4825. Since RREG is robust to outliers, the effects of outliers are down weighted using $\psi(\hat{u}_{it})$:

$$\psi(\hat{u}_{it}) = \begin{cases} \hat{u}_{it} & ; |\hat{u}_{it}| \leq b \\ b\,\text{sign}(\hat{u}_{it}) & \text{otherwise} \end{cases} \quad (9)$$

b is set to 1.345 in order to attain a 95% efficiency at the normal distribution assumption.

Corollary 1:

When b tends to infinity, (5) coincides with CD_{lm} and (6) coincides with PCD. Therefore, $RCD1$ converges to $\chi^2_{\frac{N(N-1)}{2}}$ and $RPCD1$ tends to $N(0,1)$ under the null hypothesis of no cross dependency.

3.2 RREG with Diagnostic Tool

Diagnostic tool is usually used in regression analysis. By incorporating the use of diagnostics tool, we proposed a new sets of tests as follows:

i. Robust LM Test 2

$$RCD2 = T\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \hat{\rho}'^2_{ij} \quad (10)$$

ii. Robust PCD Test 2

$$RPCD2 = \sqrt{\frac{2T}{N(N-1)}}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \hat{\rho}'_{ij} \quad (11)$$

where $$\hat{\rho}'_{ij} = \hat{\rho}'_{ji} = \frac{\sum_{t=1}^{T} \hat{e}^{\circ}_{it}\hat{e}^{\circ}_{jt}}{\left(\sum_{t=1}^{T} \hat{e}^{\circ 2}_{it}\right)^{1/2}\left(\sum_{t=1}^{T} \hat{e}^{\circ 2}_{jt}\right)^{1/2}} \quad (12)$$

$$\hat{e}^{\circ}_{it} = \hat{e}^{\circ}_{jt} = \begin{cases} \hat{e}'_{it} & ; |\hat{u}_{it}| \le d \\ 0 & \text{otherwise} \end{cases} \quad ; \quad (13)$$

$\hat{e}'_{it}$ is the fitted residual obtained from RREG, $\hat{u}_{it}$ is computed as in (8) and d is chosen from the simulations study in Table 1.

Corollary 2:

When b tends to infinity, (10) coincides with CD_{lm} and (11) coincides with PCD. Therefore, $RCD2$ converges to $\chi^2_{\frac{N(N-1)}{2}}$ and $RPCD2$ tends to $N(0,1)$ under the null hypothesis of no cross dependency.

4. Monte Carlo Simulation Study:

Consider the following data generating process (DGP);

$Y_{it} = \alpha_i + \beta_i X_{it} + e_{it}$; and

$e_{it} = \gamma_i f_t + \varepsilon_{it}; \alpha_i \sim iidU(-0.5,0.5)$; $\beta_i \sim iidN(1,0.02)$;

$X_{it} \sim iidN(0,1)$, $\varepsilon_{it} \sim iidN(0,1)$; $f_t \sim iidN(0,1)$;

Under various degrees of cross dependency, we set the value of γ_i as follows:

(i) $\gamma_i = 0$ for no cross dependency;

(ii) $\gamma_i \sim iidU(0.1,0.3)$ for mild cross dependency and ;

(iii) $\gamma_i \sim iidU(0.5,1.5)$ for strong effect of cross dependency.

In order to see how these methods perform in the presence of outliers, we locate m_{it} (contaminations) at different time, τ_i for each cross section. Let

$$e_{it} = \begin{cases} e_{it} & \text{for } t \ne \tau_i \\ e_{it} + m_{it} & \text{for } t = \tau_i \end{cases} \quad \text{for } i = 1,2,\ldots\ldots,N;$$

In this simulations study, we investigate the performance of the cross dependency test (existing and the proposed procedure) with several types of contaminations. We compute the mean and standard deviations values of the test based on 1000 replications for $N = 5, T = 5$ and 500 replications for $N = 20, T = 100$, respectively. We choose to study these samples in order to investigate the performance of the test in small size and large sample. We also conduct a simulation study to compute the level and power of the study in the presence of outliers. Fifty observations will be discarded to eliminate the initial (zero) effects of the data.

4.1 Results

Table 1: Critical region for d based on 1000 replications.

Nominal Level	0.1	0.05	0.025	0.01
$N = 5, T = 5$	2.204	3.013	3.900	5.350
$N = 20, T = 100$	1.655	1.977	2.265	2.614

Table 1 provides the results of critical region, d in (13) based on four different nominal levels. At 5% significant level, we found $d \approx 3.00$ and $d \approx 2.00$ for $N = 5, T = 5$ and $N = 20, T = 100$, respectively. The result is used in computing the tests statistics which are reported in Table 2-7.

Tables 2 - 4 display results of the cross dependency test for two different pair of sample size and various types of contaminations. The number reported is the average and standard deviation (in parentheses) of the tests over 1000 replications for $N = 5, T = 5$ and 500 replications for $N = 20, T = 100$, respectively. The performances of the tests are analyzed using size and power of the study and are reported in Table 5-7.

Table 2 provides the average of the cross dependency test with and without the presence of cross sectional dependence for , $N = 5, T = 5$, both with small size. For the case of no cross dependency $(\gamma_i = 0)$, the average and standard deviation for all methods are close to the corresponding distribution, that is CD_{lm} is $\chi^2_{(N(N-1)/2)}$, with mean 10 and standard deviation 4.47 while *PCD* is Normal with mean 0 and a unit variance , respectively. Similar results were obtained for $\gamma_i \sim iidU(0.1,0.3)$. This indicates that small effect of γ_i does not affect the innovations. When innovations are normally distributed with $\gamma_i \sim iidU(0.5,1.5)$, *PCD* and *RPCD1* give significant average of the existing correlation across units in panel. The CD_{lm} results however are about the same when no cross dependency occur in panel. The size of the test as reported in Table 5 is reasonable with all values close to 0.05. We compute the power of each test with the value of $\gamma_i \sim iidU(0.1,0.3)$ and $\gamma_i \sim iidU(0.5,1.5)$. In the presence of mild cross dependency, the powers of tests close to the size of the study for all conditions. For $\gamma_i \sim iidU(0.5,1.5)$, the proportion of rejecting the null hypothesis is less than two times in clean data in contaminated panel. The test however is still poorer in detecting cross section correlation with and without the presence of outliers. This means the proposed methods provide a comparable result as in *PCD* in small sample size.

Table 3 displays similar design of simulation study but for larger pair of sample size; $N = 20, T = 100$. When there is no contaminations in panel, the average results of test statistics for all methods are acceptable for $\gamma_i = 0$ and $\gamma_i \sim iidU(0.5,1.5)$. The size and power of the tests for strong cross dependency; i.e. $\gamma_i \sim iidU(0.5,1.5)$ (Table 6) are also reliable with all values being close to 0.05 and 1.00, respectively. CD_{lm} , *RCD1* and *RCD2*, however, provide insignificant value of average cross dependency test when $\gamma_i \sim iidU(0.1,0.3)$ with small proportion of rejection of no cross dependency as reported in Table 6. In the presence of outliers with $\gamma_i = 0$, almost all tests provide the insignificant (i.e. do not reject the null) average of the test and this is shown via the size of the test in Table 6.The proposed procedure, *RCD1* and *RCD2* provide a good size of the test and they are comparable with *PCD* and its robust version, *RPCD2*.

For $\gamma_i \sim iidU(0.1,0.3)$, *RPCD1* and *RPCD2* provide the significant value of the test over 500 replications. The power of those tests (Table 6) are reasonable with all values greater than 0.75 even for contaminated panel. The power of CD_{lm} is small and its robust version provides the similar results in contaminated panel. Thus, CD_{lm} of the proposed procedure is less powerful in rejecting the null when the effect of cross dependency is mild in panel. Neverthelss, it is however better than that of the *PCD*.

With $\gamma_i \sim iidU(0.5,1.5)$, the proposed methods (*RCD1*, *RCD2*, *RPCD1* and *RPCD2*) outperform those CD_{lm} and *PCD* in both clean and contaminated panel. The average of CD_{lm} and *PCD* failed to reject the null hypothesis in the presence of outliers generated from χ^2_{30} and Cauchy distribution. As we know, OLS is non-robust to outliers and applying this method of estimation will provide an inaccurate result of the test. Using OLS in both tests give us lack of power of study in the presence of outliers as stated in Table 6.

As we decrease the percentage of contaminations (see Table 4), the previous results have changes from rejection to not rejection in the presence of 5% outliers. The size of the test in Table 7 however does not differ the number of outliers is reduced. However, the proposed tests show a reasonable size in all situations. In the presence of cross dependency, for $\gamma_i \sim iidU(0.1,0.3)$ the power of the test remains the same for all conditions. When the effect of cross dependency increased; i.e. $\gamma_i \sim iidU(0.5,1.5)$, the power also increased except for CD_{lm}. The powers of other tests is almost unchanged with 5% contaminations.

Table 2: Cross Dependency Test for $N = 5, T = 5$

γ_i	Test / e_{it}	CD_{lm}	$RCD1$	$RCD2$	PCD	$RPCD1$	$RPCD2$
$\gamma_i = 0$	Normal	12.354	11.701	11.428	0.023	0.009	-0.023
		(3.940)	(3.820)	(3.802)	(1.115)	(1.070)	(1.042)
	$0.90\ N(0,1) + 0.10\chi^2_{30}$	12.239	11.384	11.097	0.056	0.027	0.043
		(3.578)	(3.551)	(3.664)	(1.061)	(1.035)	(1.092)
	$0.90\ N(0,1) + 0.10N(4,4)$	12.547	11.725	11.430	-0.062	-0.027	0.033
		(4.026)	(3.696)	(3.707)	(1.108)	(1.097)	(1.072)
	$0.90\ N(0,1) + 0.10$ Lognorm	12.591	11.883	11.642	-0.031	-0.006	0.042
		(4.172)	(3.829)	(3.942)	(1.103)	(1.066)	(1.063)
	$0.90\ N(0,1) + 0.10$ Cauchy	12.451	11.624	11.298	-0.025	0.004	-0.023
$\gamma_i \sim iidU(0.1,0.3)$	Normal	12.430	11.895	11.447	0.150	0.128	0.079
		(3.911)	(3.890)	(3.683)	(1.166)	(1.135)	(1.089)
	$0.90\ N(0,1) + 0.10\chi^2_{30}$	12.351	11.537	11.454	0.051	0.087	0.119
		(4.023)	(3.715)	(3.848)	(1.146)	(1.101)	(1.085)
	$0.90\ N(0,1) + 0.10N(4,4)$	12.718	11.715	11.375	0.084	0.099	0.083
		(4.235)	(3.895)	(3.981)	(1.236)	(1.153)	(1.121)
	$0.90\ N(0,1) + 0.10$ Lognorm	12.733	11.859	11.562	0.092	0.099	0.063
		(4.163)	(4.017)	(4.111)	(1.202)	(1.140)	(1.105)
	$0.90\ N(0,1) + 0.10$ Cauchy	12.489	11.542	11.471	0.091	0.078	0.045
		(3.948)	(3.707)	(3.908)	(1.183)	(1.135)	(1.119)
$\gamma_i \sim iidU(0.5,1.5)$	Normal	15.483	14.328	13.330	2.177*	2.000 *	1.601
		(5.744)	(5.373)	(5.127)	(1.674)	(1.598)	(1.610)
	$0.90\ N(0,1) + 0.10\chi^2_{30}$	13.456	12.535	12.341	0.827	1.063	1.097
		(4.706)	(4.373)	(4.699)	(1.562)	(1.515)	(1.460)
	$0.90\ N(0,1) + 0.10N(4,4)$	13.697	12.832	12.388	1.118	1.263	1.149
		(4.763)	(4.581)	(4.420)	(1.552)	(1.527)	(1.497)
	$0.90\ N(0,1) + 0.10$ Lognorm	14.136	13.313	12.508	1.455	1.516	1.301
		(4.917)	(4.837)	(4.595)	(1.614)	(1.573)	(1.507)
	$0.90\ N(0,1) + 0.10$ Cauchy	13.684	12.731	12.452	1.112	1.249	1.199
		(4.551)	(4.363)	(4.551)	(1.574)	(1.547)	(1.542)

Note: The results are the sample mean and standard deviation (parentheses) of CD_{lm} and PCD respectively, based on 1000 replications.

*The test is significant when $CD_{lm} > \chi^2_{(N(N-1)/2)} = 18.30$ and $|PCD| > N(0,1) = 1.96$ at 5% significant levels.

Table 3: Cross Sectional Dependence test for $N = 20, T = 100$ (10% contaminations)

γ_i	Test / e_{it}	CD_{lm}	$RCD1$	$RCD2$	PCD	$RPCD1$	$RPCD2$
$\gamma_i = 0$	Normal	192.030	191.808	190.944	-0.001	0.014	-0.001
		(19.234)	(19.913)	(20.511)	(1.006)	(1.030)	(1.047)
	$0.90\ N(0,1) + 0.10\chi^2_{30}$	192.392	191.904	192.448	0.050	2.763 *	-0.017
		(23.090)	(20.223)	(18.938)	(1.033)	(1.019)	(0.983)
	$0.90\ N(0,1) + 0.10N(4,4)$	188.826	191.014	190.925	0.045	0.127	0.031
		(34.273)	(17.830)	(18.589)	(0.954)	(0.981)	(1.020)
	$0.90\ N(0,1) + 0.10$ Lognorm	190.376	190.036	190.759	0.034	0.599	0.086
		(73.393)	(18.404)	(18.393)	(0.984)	(0.995)	(1.000)
	$0.90\ N(0,1) + 0.10$ Cauchy	190.769	191.465	192.289	0.036	0.000	0.048
		(80.260)	(19.966)	(18.577)	(0.942)	(1.022)	(1.029)
$\gamma_i \sim iidU(0.1,0.3)$	Normal	221.247	218.330	205.936	5.192*	4.938*	3.761*
		(27.763)	(26.219)	(22.291)	(1.717)	(1.665)	(1.526)
	$0.90\ N(0,1) + 0.10\chi^2_{30}$	190.709	225.516*	210.387	0.010	6.163*	3.077*
		(21.801)	(27.645)	(20.563)	(0.978)	(1.544)	(1.423)
	$0.90\ N(0,1) + 0.10N(4,4)$	191.558	208.878	207.556	0.202	3.657*	3.730*
		(31.546)	(22.800)	(23.079)	(0.984)	(1.445)	(1.536)
	$0.90\ N(0,1) + 0.10$ Lognorm	185.986	213.742	204.671	0.228	4.524*	3.553*
		(71.183)	(25.387)	(22.301)	(0.981)	(1.570)	(1.505)
	$0.90\ N(0,1) + 0.10$ Cauchy	191.700	207.921	208.190	0.087	3.569*	3.734*
		(78.097)	(22.289)	(23.219)	(1.025)	(1.491)	(1.529)
$\gamma_i \sim iidU(0.5,1.5)$	Normal	4393.100*	4040.739*	2511.355*	63.789*	61.087*	47.806 *
		(792.487)	(763.288)	(608.222)	(6.346)	(6.340)	(6.301)
	$0.90\ N(0,1) + 0.10\chi^2_{30}$	193.285	2315.331*	1762.913 *	1.520	45.287*	39.385*
		(23.379)	(464.728)	(454.738)	(1.179)	(5.121)	(5.770)
	$0.90\ N(0,1) + 0.10N(4,4)$	223.171*	2247.844*	2251.347*	5.562*	44.305*	45.008*
		(41.805)	(474.317)	(535.675)	(1.977)	(5.321)	(5.988)
	$0.90\ N(0,1) + 0.10$ Lognorm	281.515 *	2845.936*	2112.504*	5.556*	50.497*	43.469*
		(108.271)	(597.721)	(529.529)	(2.773)	(5.977)	(6.092)
	$0.90\ N(0,1) + 0.10$ Cauchy	194.066	2192.461*	2345.785*	0.804	43.680*	45.950*
		(77.414)	(464.100)	(559.377)	(1.183)	(5.299)	(6.098)

Note: The results are the sample mean and standard deviation (parentheses) of CD_{lm} and PCD respectively, based on 500 replications.

*The test is significant when $CD_{lm} > \chi^2_{(N(N-1)/2)} = 223.160$ and $|PCD| > N(0,1) = 1.96$ at 5% significant levels.

Table 4: Cross Sectional Dependence test for $N = 20, T = 100$ (5% contaminations)

γ_i	Test e_{it}	CD_{lm}	$RCD1$	$RCD2$	PCD	$RPCD1$	$RPCD2$
$\gamma_i = 0$	Normal	192.030	191.808	190.944	-0.001	0.014	-0.001
		(19.234)	(19.913)	(20.511)	(1.006)	(1.030)	(1.047)
	$0.95\ N(0,1) + 0.05\chi^2_{30}$	192.581	189.398	192.186	-0.022	0.733	-0.013
		(38.275)	(19.494)	(19.054)	(1.031)	(1.009)	(0.979)
	$0.95\ N(0,1) + 0.05N(4,4)$	191.581	191.385	190.513	0.019	0.058	0.026
		(45.886)	(17.848)	(18.298)	(1.008)	(1.001)	(1.010)
	$0.95\ N(0,1) + 0.05$ Lognorm	188.421	190.953	191.287	0.030	0.202	0.018
		(63.636)	(18.735)	(18.994)	(1.008)	(1.035)	(1.008)
	$0.95\ N(0,1) + 0.05$ Cauchy	185.691	191.025	192.215	-0.038	-0.003	0.011
		(76.961)	(19.462)	(18.619)	(1.003)	(1.022)	(1.046)
$\gamma_i \sim iidU(0.1,0.3)$	Normal	221.247	218.330	205.936	5.192*	4.938*	3.761*
		(27.763)	(26.219)	(22.291)	(1.717)	(1.665)	(1.526)
	$0.95\ N(0,1) + 0.05\chi^2_{30}$	188.415	215.971	203.596	0.109	4.789*	3.369*
		(36.037)	(27.004)	(22.344)	(1.042)	(1.645)	(1.510)
	$0.95\ N(0,1) + 0.05N(4,4)$	190.858	214.518	207.463	0.592	4.287*	3.827*
		(43.828)	(24.075)	(22.736)	(0.994)	(1.564)	(1.541)
	$0.95\ N(0,1) + 0.05$ Lognorm	189.204	216.534	206.771	0.898	4.634*	3.723*
		(59.166)	(24.408)	(22.798)	(1.097)	(1.621)	(1.523)
	$0.95\ N(0,1) + 0.05$ Cauchy	194.555	213.401	208.202	0.211	4.263*	3.744*
		(84.603)	(25.076)	(23.160)	(1.016)	(1.631)	(1.581)
$\gamma_i \sim iidU(0.5,1.5)$	Normal	4393.100*	4040.739*	2511.355*	63.789*	61.087*	47.806*
		(792.487)	(763.288)	(608.222)	(6.346)	(6.340)	(6.301)
	$0.95\ N(0,1) + 0.05\chi^2_{30}$	205.248	2925.314*	2088.451*	3.368*	51.295*	43.234*
		(41.902)	(598.863)	(523.827)	(1.587)	(5.858)	(6.034)
	$0.95\ N(0,1) + 0.05N(4,4)$	403.501*	3044.824*	2414.057*	12.095*	52.355*	46.776*
		(122.032)	(632.519)	(599.387)	(3.313)	(6.087)	(6.403)
	$0.95\ N(0,1) + 0.05$ Lognorm	712.852*	3398.364 *	2321.304*	15.738*	55.584*	45.778*
		(274.245)	(706.001)	(595.814)	(5.112)	(6.427)	(6.468)
	$0.95\ N(0,1) + 0.05$ Cauchy	239.534*	3008.858*	2452.455*	3.479*	52.026*	47.111*
		(95.069)	(630.438)	(611.719)	(2.222)	(6.097)	(6.482)

Note: The results are the sample mean and standard deviation (parentheses) of CD_{lm} and PCD respectively, based on 500 replications.

*The test is significant when $CD_{lm} > \chi^2_{(N(N-1)/2)} = 223.160$ and $|PCD| > N(0,1) = 1.96$ at 5% significant levels.

Table 5: Size and power of study under the null hypothesis of no cross sectional dependence for $N = 5, T = 5$

γ_i	Test e_{it}	CD_{lm}	$RCD1$	$RCD2$	PCD	$RPCD1$	$RPCD2$
Size							
$\gamma_i = 0$	Normal	0.077	0.049	0.052	0.065	0.052	0.047
	$0.90\ N(0,1) + 0.10\chi^2_{30}$	0.059	0.042	0.029	0.058	0.052	0.049
	$0.90\ N(0,1) + 0.10N(4,4)$	0.079	0.052	0.043	0.055	0.054	0.061
	$0.90\ N(0,1) + 0.10$ Lognorm	0.086	0.057	0.054	0.057	0.051	0.052
	$0.90\ N(0,1) + 0.10$ Cauchy	0.074	0.049	0.041	0.056	0.052	0.043
Power							
$\gamma_i \sim iidU(0.1,0.3)$	Normal	0.081	0.066	0.044	0.079	0.071	0.057
	$0.90\ N(0,1) + 0.10\chi^2_{30}$	0.074	0.049	0.040	0.073	0.070	0.052
	$0.90\ N(0,1) + 0.10N(4,4)$	0.105	0.062	0.051	0.093	0.078	0.061
	$0.90\ N(0,1) + 0.10$ Lognorm	0.101	0.066	0.060	0.076	0.068	0.059
	$0.90\ N(0,1) + 0.10$ Cauchy	0.087	0.059	0.043	0.077	0.070	0.064
$\gamma_i \sim iidU(0.5,1.5)$	Normal	0.256	0.200	0.150	0.547	0.500	0.401
	$0.90\ N(0,1) + 0.10\chi^2_{30}$	0.133	0.102	0.179	0.235	0.278	0.228
	$0.90\ N(0,1) + 0.10N(4,4)$	0.146	0.111	0.092	0.294	0.328	0.289
	$0.90\ N(0,1) + 0.10$ Lognorm	0.158	0.135	0.106	0.377	0.384	0.316
	$0.90\ N(0,1) + 0.10$ Cauchy	0.139	0.100	0.088	0.293	0.319	0.268

Note: Reject Ho when $CD_{lm} > \chi^2_{(N(N-1)/2)} = 18.30$ and $|PCD| > N(0,1) = 1.96$ 5% significant levels.

Results are computed based on 1000 replications.

Table 6: Size and power of study under the null hypothesis of no cross sectional dependence for $N = 20, T = 100$ (10% contaminations)

	Test e_{it}	CD_{lm}	$RCD1$	$RCD2$	PCD	$RPCD1$	$RPCD2$
Size							
	Normal	0.056	0.046	0.050	0.042	0.066	0.062
	0.90 $N(0,1)+0.10\chi^2_{30}$	0.104	0.062	0.056	0.054	0.774	0.046
$\gamma_i = 0$	0.90 $N(0,1)+0.10N(4,4)$	0.154	0.038	0.046	0.040	0.060	0.064
	0.90 $N(0,1)+0.10$ Lognorm	0.282	0.036	0.048	0.044	0.086	0.054
	0.90 $N(0,1)+0.10$ Cauchy	0.290	0.068	0.054	0.032	0.052	0.058
Power							
	Normal	0.428	0.380	0.212	0.986	0.972	0.884
	0.90 $N(0,1)+0.10\chi^2_{30}$	0.074	0.490	0.138	0.046	1.000	0.760
$\gamma_i \sim iidU(0.1,0.3)$	0.90 $N(0,1)+0.10N(4,4)$	0.144	0.236	0.236	0.062	0.886	0.872
	0.90 $N(0,1)+0.10$ Lognorm	0.256	0.332	0.182	0.056	0.962	0.850
	0.90 $N(0,1)+0.10$ Cauchy	0.300	0.228	0.242	0.050	0.864	0.852
	Normal	1.000	1.000	1.000	1.000	1.000	1.000
	0.90 $N(0,1)+0.10\chi^2_{30}$	0.098	1.000	1.000	0.346	1.000	1.000
$\gamma_i \sim iidU(0.5,1.5)$	0.90 $N(0,1)+0.10N(4,4)$	0.564	1.000	1.000	0.970	1.000	1.000
	0.90 $N(0,1)+0.10$ Lognorm	0.686	1.000	1.000	0.924	1.000	1.000
	0.90 $N(0,1)+0.10$ Cauchy	0.322	1.000	1.000	0.170	1.000	1.000

Note: Reject Ho when $CD_{lm} > \chi^2_{(N(N-1)/2)} = 223.160$ and $|PCD| > N(0,1) = 1.96$ at 5% significant levels.
Results are computed based on 500 replications.

Table 7: Size and power of study under the null hypothesis of no cross sectional dependence for $N = 20, T = 100$ (5% contaminations)

	Test e_{it}	CD_{lm}	$RCD1$	$RCD2$	PCD	$RPCD1$	$RPCD2$
Size							
	Normal	0.056	0.046	0.050	0.042	0.066	0.062
	0.95 $N(0,1)+0.05\chi^2_{30}$	0.180	0.048	0.054	0.054	0.134	0.042
$\gamma_i = 0$	0.95 $N(0,1)+0.05N(4,4)$	0.242	0.042	0.040	0.066	0.052	0.058
	0.95 $N(0,1)+0.05$ Lognorm	0.252	0.042	0.048	0.052	0.070	0.050
	0.95 $N(0,1)+0.05$ Cauchy	0.272	0.066	0.050	0.052	0.056	0.056
Power							
	Normal	0.428	0.380	0.212	0.986	0.972	0.884
	0.95 $N(0,1)+0.05\chi^2_{30}$	0.162	0.354	0.176	0.066	0.970	0.822
$\gamma_i \sim iidU(0.1,0.3)$	0.95 $N(0,1)+0.05N(4,4)$	0.226	0.326	0.216	0.096	0.944	0.898
	0.95 $N(0,1)+0.05$ Lognorm	0.244	0.346	0.212	0.166	0.968	0.878
	0.95 $N(0,1)+0.05$ Cauchy	0.328	0.306	0.250	0.046	0.926	0.872
	Normal	1.000	1.000	1.000	1.000	1.000	1.000
	0.95 $N(0,1)+0.05\chi^2_{30}$	0.270	1.000	1.000	0.798	1.000	1.000
$\gamma_i \sim iidU(0.5,1.5)$	0.95 $N(0,1)+0.05N(4,4)$	0.980	1.000	1.000	1.000	1.000	1.000
	0.95 $N(0,1)+0.05$ Lognorm	0.994	1.000	1.000	1.000	1.000	1.000
	0.95 $N(0,1)+0.05$ Cauchy	0.548	1.000	1.000	0.752	1.000	1.000

Note: Reject Ho when $CD_{lm} > \chi^2_{(N(N-1)/2)} = 223.160$ and $|PCD| > N(0,1) = 1.96$ at 5% significant levels.
Results are computed based on 500 replications.

6. Conclusion

The existing test for checking the existence of cross dependency among innovations using the traditional approach named OLS is highly sensitive with the presence of outliers. To overcome that problem, the robust version of them is proposed in order to capture the presence of cross dependency and spurious observations in panel data. We introduced two types of procedure on both Lagrange Multiplier of Breusch and Pagan (1980) and Pesaran's (2004) Cross Dependency Test which are based on Huber filter function and diagnostic procedure. Based on the Monte Carlo simulation study, our proposed procedure named *RPCD2* using RREG with diagnostic tool outperform *PCD* and other approaches in contaminated panel and under various degree of cross dependency. Others proposed procedure however are still able to produce reliable result and robust to outliers in the presence of strong effect of cross dependency among innovations.

References

- Bai, J. and S. Ng. 2002. Determining the Number of Factors in Approximate Factor Model. Econometrica 70. 191-221.
- Beck, N. and J. N. Katz. 1995. What to do (and not to do) with Time-Series Cross-Section Data. The American Political Science Review, Vol. 89, No. 3, (Sep., 1995), pp. 634-647
- Bruesch, T. S. and A. R. Pagan. 1980. The Lagrange Multiplier Test and Its applications to Model Specifications in Econometrics. Review of Economic Studies. XLVII. 239-253.
- Cerrato, M. 2001. The Cross Sectional Dependence Puzzle.
- Coakley, J., A. Fuertes and R. P. Smith. 2002. A Principal Components Approach to Cross-Section Dependence in Panels. Mimeo, Birkberk College, University of London.
- Coakley, J., A. M. Fuertes and R. Smith. 2006. Unobserved Heterogeneity in Panel Time Series Models. Computational Statistics and Data Analysis. 50. 2361-2380.

- Frees, E. W. 1995. Assessing Cross-sectional Correlation in Panel Data. Journal of Econometrics. 69. 393-414.
- 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.
- Hoyos, R. E. D. and V. Sarafidis. (2006). Testing for Cross-sectional Dependence in Panel Data Model. The Stata Journal. 6 (4). 482-496.
- Kapetanios, G. and M. H. Pesaran. 2004. Alternative Approaches to Estimation and Inference in Large Multifactor Panels: Small sample Results with an Application to Modeling of Asset Returns. CESIFO Working Paper.
- Moon, H. R. and B. Perron. 2004. Testing For a Unit Root in Panels with Dynamic Factors. Journal of Econometrics 122. 51-126.
- Noman, A. 2008. Testing for PPP in the Mean-Group Panel Regression Framework: Further Evidence.
- Pesaran, M.H. 2004. General Diagnostic Tests for Cross Section Dependence in Panels.
- Pesaran, M.H. 2006. Estimation and Inference in Large Heterogeneous Panels with a Multifactor Error Structure. Econometrica 74(4). 967-1012.
- Pesaran, M. H. , A. Ullah and T. Yamagata. 2008. A-Bias Adjusted LM Test of Error Cross Section Independence. The Econometrics Journal. 11(1). 105-127.
- Philips, P.C.B. and D. Sul. 2003. Dynamic Panel Estimation and Homogeneity Testing Under Cross Section Dependence. The Econometrics Journal. 6. 217-259.
- Sarafidis, V., T. Yamagata and D. Robertson. 2009. A Test of Cross Section Dependence for a Linear Dynamic Panel. Journal of Econometrics. 148. 149-161.