Capturing Cross-Sectional Correlation with Time Series: with an Application to Unit Root Test

Chor-yiu SIN (CY)
Wang Yanan Institute for Studies in Economics (WISE)
Xiamen University, Fujian, P.R.China
Email: cysin@xmu.edu.cn

Friday, May 9, 2007

China Center for Economic Research
Peking University

Preliminary draft. Comments are much appreciated.

INTRODUCTION

• Throughout the paper, we consider the following linear regression model:

$$y_{it} = x'_{it}\beta + u_{it}, \tag{1}$$

where $i = 1, \ldots, N$ and $t = 1, \ldots, T$, $T \geq 2$, x_{it} is a $kx1$-vector while both y_{it} and u_{it} are scalars.

• In fact, this is a typical *panel data* model.

• Though, as one can tell from the title, our *focus* is the time series properties.

• More precisely, we are interested, with panel data, in investigating the time series properties, with a **low** time-series dimension (T is fixed) but a **high** cross-sectional dimension ($N \rightarrow \infty$).

INTRODUCTION (cont′d)

• On major drawback in making inference on the parameter β in Equation (1) is to *model* and *estimate* the cross-sectional correlations.

• More precisely, for statistical inference, one may need to model and estimate, for $t = 1, \ldots, T$, the following $N(N-1)/2$ cross-product terms:

$$E[x_{it} u_{it} u_{jt} x'_{jt}], \tag{2}$$

where $i < j$, and $i, j = 1, \ldots, N$.

• This is not easy when N, the number of cross-sectional units, is large.

INTRODUCTION (cont′d)

• In the literature, there are at least *four* ways to tackle this issue.

• (i) Assuming away the cross-sectional correlations. That is, in Equation (2) above:

$$E[x_{it}u_{it}u_{jt}x'_{jt}] = 0.$$

• See, for instance, Anderson (1978) *JASA*, Anderson and Hsiao (1981) *JASA*, Holz-Eaken, Newey and Rosen (1988) *Ec.*, Quah (1994) *EL*, and Phillips and Moon (1999) *Ec.*

• This assumption may not be justifiable.

INTRODUCTION (cont′d)

• (ii) Assuming T, the number of time-series units, is also large. In one way or the other, one may estimate the $N(N-1)/2$ cross-product terms in Equation (2) with T time-series units.

• See, for instance, Kao (1999) *JOE*, Bai and Ng (2003) *Ec.* and Bai (2003) *Ec.* and a survey paper by Choi (2005).

• The assumption of large T is justifiable in many cases but it may not be justifiable in the so-called *short* panel data.

INTRODUCTION (cont′d)

- (iii) Using the *geographical* distance or the *economic* distance to model the cross-sectional correlations.

- Geographical distance is commonly used in the field of *spatial statistics/econometrics*. See, for instance, Kelejian and Prucha (1999) *IER*.

- The interesting idea of economic distance is first introduced by Conley (1999) *JOE*. Since then it attracts a lot of attention from economists.

- However, the concept *geographical* distance may not be applicable to some if not all economic data while the concept *economic* distance is a bit controversial.

INTRODUCTION (cont′d)

- (iv) Pesaran (2006) *Ec.*: A special case.

INTRODUCTION (cont′d)

- In this paper, we first follow the lines in Conley (1999) and prove the $\sqrt{N}$- consistency of our OLS estimator.

- Then we use the T time-series unit to *capture* the cross-sectional correlations. T can be small as long as $T \geq 2$.

- In fact, for sake of theoretical simplicity, we assume that T is fixed while $N \to \infty$.

OUTLINE OF THE TALK

(1) Introduction

(2) Two OLS estimators and two Wald tests
• The *disjoint* case (DISJ)
• The *overlapping* case (OVER) and its similarity with the classical *z-test*

(3) Two applications: Testing for unit root and testing for cointegration

(4) Generalizing and extending (2)

(5) Simulating the critical values

(6) Monte Carlo Experiments
• Comparing DISJ and OVER with another test that ignores cross-sectional correlations

(7) Conclusions and Discussions

OLS : DISJ

• For the disjoint case, we split the time-series units into two parts, one with T_1 observations and the other with $T - T_1$ observations.

• The T_1 observations are for estimating β while the remaining $T - T_1$ observations are for estimating the "variance-covariance" matrix of $\widehat{\beta}$.

• More precisely:

$$\widehat{\beta} = (\sum_{s=1}^{T_1} \sum_{i=1}^{N} x_{is} x_{is}')^{-1} (\sum_{s=1}^{T_1} \sum_{i=1}^{N} x_{is} y_{is}). \qquad (3)$$

Note the time-series units go from 1 to T_1 only.

Assumptions : DISJ

Assumption (a). $N \to \infty$ and T is fixed.

Assumption (b). For $t = 1, \ldots, T$,

$$N^{-1/2} \sum_{i=1}^{N} x_{it} u_{it} \longrightarrow_{\mathcal{L}} \Gamma W_t^k,$$

where Γ is a positive definite matrix and W_t^k is a $k - dimensional$ standard normal random vector.

Assumption (c). For $t = 1, \ldots, T$,

$$N^{-1} \sum_{i=1}^{N} x_{it} x'_{it} \to M a.s.,$$

where M is an kxk- invertible constant matrix.

Theorem : DISJ

Theorem 2.1. Suppose Assumptions (a)-(c) hold.

$$\sqrt{N}(\widehat{\beta} - \beta) \longrightarrow_{\mathcal{L}} M^{-1}\Gamma(\frac{1}{T_1}\sum_{s=1}^{T_1} W_s^k). \qquad (4)$$

• The proof of Theorem 2.1 follows the lines in Conley (1999). In fact Conley (1999) gives us some *primitive* assumptions to assume Assumption (b). The difference is on the "variance-covariance" matrix:

$$\begin{aligned}
\widehat{V} &= \widehat{A}^{-1}\widehat{B}\widehat{A}^{-1}, \\
\widehat{A} &= N^{-1}\sum_{s=1}^{T_1}\sum_{i=1}^{N} x_{is}x'_{is} \\
\widehat{B} &= \sum_{t=T_1+1}^{T}(N^{-1/2}\sum_{i=1}^{N} x_{it}\widehat{u}_{it})(N^{-1/2}\sum_{i=1}^{N} x_{it}\widehat{u}_{it})'.
\end{aligned}$$

Wald Test : DISJ

Assumption (d).
$\sum_{t=T_1+1}^{T}(W_t^k - \frac{1}{T_1}\sum_{s=1}^{T_1} W_s^k)(W_t^{k'} - \frac{1}{T_1}\sum_{s=1}^{T_1} W_s^{k'})$ is p.d. a.s.

• Assumption (d) is non-trivial. Consider the simple case that $T_1 = T_2 = 1$. If $W_1^k = W_2^k$ a.s., the term $(W_t^k - \frac{1}{T_1}\sum_{s=1}^{T_1} W_s^k)$ is identically zero a.s.

• The Wald test for $\beta = \beta_0$:

$$\widehat{\mathcal{W}} = \sqrt{N}(\widehat{\beta} - \beta_0)'\widehat{V}^{-1}\sqrt{N}(\widehat{\beta} - \beta_0),$$

Theorem 2.2. Suppose Assumptions (a)-(d) hold. $\widehat{\mathcal{W}}$ converges in distribution to:

$$\sum_{s=1}^{T_1} W_s^{k'}[\sum_{t=T_1+1}^{T}(W_t^k - \frac{1}{T_1}\sum_{s=1}^{T_1} W_s^k)(W_t^{k'} - \frac{1}{T_1}\sum_{s=1}^{T_1} W_s^{k'})]^{-1}\sum_{s=1}^{T_1} W_s^k. \tag{5}$$

OLS : OVER

- For the overlapping case, we use the all T observations are for both estimating β and estimating the "variance-covariance" matrix of $\widehat{\beta}$.

- More precisely:

$$\widehat{\beta} = (\sum_{s=1}^{T} \sum_{i=1}^{N} x_{is}x_{is}')^{-1}(\sum_{s=1}^{T} \sum_{i=1}^{N} x_{is}y_{is}). \quad (6)$$

Note the time-series units go from 1 to T.

Theorem : OVER

Theorem 2.1'. Suppose Assumptions (a)-(c) hold (as in Theorem 2.1).

$$\sqrt{N}(\widehat{\beta} - \beta) \longrightarrow_{\mathcal{L}} M^{-1}\Gamma(\frac{1}{T}\sum_{s=1}^{T} W_s^k). \qquad (7)$$

- The "variance-covariance" matrix:

$$\begin{aligned} \widehat{V} &= \widehat{A}^{-1}\widehat{B}\widehat{A}^{-1}, \\ \widehat{A} &= N^{-1}\sum_{s=1}^{T}\sum_{i=1}^{N} x_{is}x_{is}' \\ \widehat{B} &= \sum_{t=1}^{T}(N^{-1/2}\sum_{i=1}^{N} x_{it}\widehat{u}_{it})(N^{-1/2}\sum_{i=1}^{N} x_{it}\widehat{u}_{it})'. \end{aligned}$$

Wald Test : OVER

Assumption (d').
$\sum_{t=1}^{T}(W_t^k - \frac{1}{T}\sum_{s=1}^{T} W_s^k)(W_t^{k'} - \frac{1}{T}\sum_{s=1}^{T} W_s^{k'})$ is p.d. a.s.

- The Wald test for $\beta = \beta_0$:

$$\widehat{\mathcal{W}} = \sqrt{N}(\widehat{\beta} - \beta_0)'\widehat{V}^{-1}\sqrt{N}(\widehat{\beta} - \beta_0),$$

Theorem 2.2'. Suppose Assumptions (a)-(d) hold. $\widehat{\mathcal{W}}$ converges in distribution to:

$$\sum_{s=1}^{T} W_s^{k'}[\sum_{t=1}^{T}(W_t^k - \frac{1}{T}\sum_{s=1}^{T} W_s^k)(W_t^{k'} - \frac{1}{T}\sum_{s=1}^{T} W_s^{k'})]^{-1}\sum_{s=1}^{T} W_s^k. \quad (8)$$

- Remarks:

(i) It is not difficult to generalize the Wald tests to the case that $H_0 : R\beta = r_0$.
(ii) The distribution in Theorem 2.2' is obviously different from that in Theorem 2.2. Both of them can be simulated though.

OVER vs $z - test$

- Our OVER is analogous to the classical *z-test* for the population mean.

- Consider a special case in Equation (1):

$$y_{it} = \beta + u_{it}. \tag{9}$$

- Suppose we want to test $H_0 : \sqrt{N}\beta = \sqrt{N}\beta_0$.

- If we sum all the terms in Equation (9) against i and multiply them by $N^{-1/2}$, we will get:

$$v_{Nt} = \sqrt{N}\beta + N^{-1/2} \sum_{i=1}^{N} u_{it}, \tag{10}$$

where $v_{Nt} \equiv N^{-1/2} \sum_{i=1}^{N} y_{it}$.

OVER vs $z - test$

- Our OVER in Theorem (2.1') will give:

$$\begin{aligned} & \frac{\sqrt{T}(\bar{v}_N - \sqrt{N}\beta_0)}{\sqrt{\sum_{t=1}^T (v_{Nt} - \bar{v}_N)^2}} \\ = & \sqrt{\frac{T}{T-1}} \frac{(\bar{v}_N - \sqrt{N}\beta_0)}{\sqrt{\sum_{t=1}^T (v_{Nt} - \bar{v}_N)^2/(T-1)}} \\ \longrightarrow_{\mathcal{L}} & \sqrt{\frac{T}{T-1}} z_{T-1}, \end{aligned} \tag{11}$$

where z_{T-1} denotes a random variable which is t distributed with $T-1$ degrees of freedom.

Application : Unit Root Test

• Assuming an AR(k+1) model, we consider the linear regression model:

$$\triangle w_{it} = x'_{it}\beta + u_{it}, \tag{12}$$

where $x_{it} = (w_{it-1}, \triangle w_{it-1}, \ldots, \triangle w_{it-k+1})'$, $t = 1, \ldots, T$ and $i = 1, \ldots, N$.

• The Augmented Dickey-Fuller test in this setting is simply testing $H_0 : \beta_1 = 0$.

Application : Cointegration Test

- Presumably all the elements of w_{it} are $I(1)$. We consider the following linear regression model:

$$w_{it0} = x_{it}'\beta + u_{it}, \tag{13}$$

where $x_{it} = (w_{it1}, \ldots, w_{itk})'$, $t = 1, \ldots, T$ and $i = 1, \ldots, N$.

- One form of testing for *no* cointegration can be cast as $H_0 : \beta = 0$.

- There should not be a problem of "spurious regression" (see Granger and Newbold (1973) *JOE* and Phillips (1986) *JOE*) as we assume T is fixed.

Generalization of OLS

- Define $\mathcal{T} \equiv \{1, \ldots, T\}$. Consider two subsets of $\mathcal{T}$, $\mathcal{T}_1$ and $\mathcal{T}_2$.

- Consider the general version of OLS:

$$\hat{\hat{\beta}} = (\sum_{s \in \mathcal{T}_1} \sum_{i=1}^{N} x_{is} x_{is}')^{-1} (\sum_{s \in \mathcal{T}_1} \sum_{i=1}^{N} x_{is} y_{is}). \quad (14)$$

Theorem 4.1. Suppose Assumptions (a)-(c) hold.

$$\sqrt{N}(\hat{\hat{\beta}} - \beta) \longrightarrow_{\mathcal{L}} M^{-1} \Gamma (\frac{1}{\# \mathcal{T}_1} \sum_{s \in \mathcal{T}_1} W_s^k).$$

- $\hat{V}$ can be defined accordingly, with the time-series observations in the subset $\mathcal{T}_2$,

- The Wald test can also be constructed accordingly.

Extenstion to Instrumental Variable Estimation

- Define $\mathcal{T} \equiv \{1, \ldots, T\}$. Consider two subsets of $\mathcal{T}$, $\mathcal{T}_1$ and $\mathcal{T}_2$.

- Suppose we have an instrument z_{it}, which is also a $kx1$-vector. Define the following IV (instrumental variable) estimator:

$$\tilde{\beta} = (\sum_{s \in \mathcal{T}_1} \sum_{i=1}^{N} z_{is} x'_{is})^{-1} (\sum_{s \in \mathcal{T}_1} \sum_{i=1}^{N} z_{is} y_{is}). \quad (15)$$

- **Assumption (b')**. For $t = 1, \ldots, T$,

$$N^{-1/2} \sum_{i=1}^{N} z_{it} u_{it} \longrightarrow_{\mathcal{L}} \Gamma W_t^k,$$

where Γ is a positive definite matrix and W_t^k is a $k - dimensional$ standard normal random vector.

Extenstion to Instrumental Variable Estimation

- **Assumption (c')**. For $t = 1, \ldots, T$,

$$N^{-1} \sum_{i=1}^{N} z_{it} x'_{it} \to M a.s.,$$

where M is an kxk- invertible constant matrix.

- **Theorem 4.3**. Suppose Assumptions (a), and Assumptions (b')-(c') hold.

$$\sqrt{N}(\tilde{\beta} - \beta) \longrightarrow_{\mathcal{L}} M^{-1} \Gamma(\frac{1}{\# \mathcal{T}_1} \sum_{s \in \mathcal{T}_1} W_s^k).$$

- $\tilde{V}$ can be defined accordingly, with the time-series observations in the subset $\mathcal{T}_2$,

- The Wald test can also be constructed accordingly.

Simulating Critical Values

TABLE 5.1

Quantiles of the Limiting Distribution in (5) or (8), k = 1.

		$\alpha-$th simulated quantiles				
T	rv	.800	.900	.950	.980	.990
2	$DISJ$	2.806	10.502	40.500	267.384	1063.563
	$OVER$	18.948	79.502	320.144	2118.335	8564.449
	$2z_1^2$	18.948	79.733	322.885	2025.152	8104.427
3	$DISJ$	9.273	36.517	147.250	947.310	3452.401
	$OVER$	5.375	12.882	27.866	74.468	151.616
	$\frac{3}{2}z_2^2$	5.335	12.790	27.774	72.767	147.758
4	$DISJ$	1.775	3.593	7.110	17.652	35.444
	$OVER$	3.579	7.386	13.491	27.004	44.591
	$\frac{4}{3}z_3^2$	3.577	7.382	13.500	27.494	45.490
5	$DISJ$	3.225	6.918	14.079	34.709	70.060
	$OVER$	2.927	5.680	9.597	17.531	26.328
	$\frac{5}{4}z_4^2$	2.938	5.682	9.633	17.550	26.496
6	$DISJ$	1.639	2.906	4.834	9.117	14.410
	$OVER$	2.625	4.853	7.914	13.631	19.782
	$\frac{6}{5}z_5^2$	2.614	4.872	7.932	13.588	19.508
7	$DISJ$	2.423	4.411	7.426	14.067	22.642
	$OVER$	2.412	4.410	6.974	11.525	15.841
	$\frac{7}{6}z_6^2$	2.419	4.404	6.986	11.525	16.032
8	$DISJ$	1.627	2.719	4.119	7.006	10.005
	$OVER$	2.233	3.978	6.058	10.145	13.845
	$\frac{8}{7}z_7^2$	2.288	4.104	6.392	10.272	13.992
9	$DISJ$	2.156	3.693	5.742	9.869	15.029
	$OVER$	2.177	3.827	5.812	9.132	12.295
	$\frac{9}{8}z_8^2$	2.195	3.892	5.982	8.858	12.663
10	$DISJ$	1.615	2.637	3.957	6.100	8.419
	$OVER$	2.090	3.694	5.617	8.636	11.601
	$\frac{10}{9}z_9^2$	2.125	3.733	5.685	8.842	11.736
20	$DISJ$	1.614	2.570	3.607	5.020	6.104
	$OVER$	1.835	3.073	4.518	6.654	8.273
	$\frac{20}{19}z_{19}^2$	1.856	3.147	4.611	6.786	8.616
30	$DISJ$	1.600	2.560	3.577	4.964	6.215
	$OVER$	1.761	2.963	4.232	6.082	7.589
	$\frac{30}{29}z_{29}^2$	1.778	2.986	4.326	6.270	7.857
50	$DISJ$	1.602	2.572	3.676	5.066	6.084
	$OVER$	1.715	2.854	4.108	5.852	7.247
	$\frac{50}{49}z_{49}^2$	1.722	2.868	4.121	5.902	7.329
100	$DISJ$	1.627	2.643	3.732	5.208	6.266
	$OVER$	1.667	2.782	3.981	5.692	6.824
	$\frac{100}{99}z_{99}^2$	1.681	2.785	3.977	5.648	6.968
121	$DISJ$	1.643	2.721	3.846	5.249	6.173
	$OVER$	1.670	2.792	4.023	5.702	7.004
	$\frac{121}{120}z_{120}^2$	1.652	2.772	3.953	5.606	6.906
χ_1^2		1.642	2.706	3.841	5.412	6.635

Monte – Carlo Experiements

TABLE 6.1(a)

Rejection Percentage under $H_0 : \beta_1 = 0$, $\rho = 0$.

		Size		
T	$Test$	10%	5%	1%
2	$DISJ$	10.00	4.65	0.75
	$OVER$	9.85	4.65	0.70
	$WHITE$	11.45	6.95	1.75
10	$DISJ$	10.00	5.05	1.00
	$OVER$	10.15	4.80	0.90
	$WHITE$	10.65	4.80	1.25
50	$DISJ$	10.45	5.15	0.90
	$OVER$	10.05	4.25	1.05
	$WHITE$	9.95	5.15	0.75

TABLE 6.1(b)

Rejection Percentage under $H_0 : \beta_1 = 0$, $\rho = 0.5$.

		Size		
T	$Test$	10%	5%	1%
2	$DISJ$	9.75	5.60	1.35
	$OVER$	9.65	5.35	1.35
	$WHITE$	22.35	15.85	5.80
10	$DISJ$	10.80	5.65	1.05
	$OVER$	10.40	5.60	1.55
	$WHITE$	20.65	13.20	4.55
50	$DISJ$	11.25	5.75	1.45
	$OVER$	11.30	5.20	1.25
	$WHITE$	20.80	13.55	4.85

TABLE 6.1(c)

Rejection Percentage under $H_0 : \beta_1 = 0$, $\rho = 0.9$.

		Size		
T	$Test$	10%	5%	1%
2	$DISJ$	12.45	6.45	1.10
	$OVER$	11.85	6.05	1.00
	$WHITE$	62.05	55.45	44.30
10	$DISJ$	12.10	7.40	2.25
	$OVER$	11.35	5.85	1.65
	$WHITE$	58.00	51.70	39.60
50	$DISJ$	11.05	6.10	2.10
	$OVER$	10.75	5.80	1.35
	$WHITE$	57.30	49.70	36.70

TABLE 6.2(a)

Rejection Percentage under $H_a : \beta_1 = 0.1,\ \rho = 0.$

T	$Test$	Size 10%	5%	1%
2	$DISJ$	15.10	7.35	1.35
	$OVER$	14.60	7.15	1.40
	$WHITE$	29.50	20.00	7.80
10	$DISJ$	49.25	37.65	14.95
	$OVER$	67.70	52.80	24.35
	$WHITE$	73.50	62.65	39.40
50	$DISJ$	95.70	92.90	83.75
	$OVER$	99.95	99.90	98.65
	$WHITE$	99.95	99.95	99.20

TABLE 6.2(b)

Rejection Percentage under $H_a : \beta_1 = 0.5,\ \rho = 0.$

T	$Test$	Size 10%	5%	1%
2	$DISJ$	56.05	31.45	5.60
	$OVER$	57.20	31.15	5.45
	$WHITE$	99.95	99.65	99.05
10	$DISJ$	100.00	100.00	99.50
	$OVER$	100.00	100.00	100.00
	$WHITE$	100.00	100.00	100.00
50	$DISJ$	100.00	100.00	100.00
	$OVER$	100.00	100.00	100.00
	$WHITE$	100.00	100.00	100.00

TABLE 6.2(c)

Rejection Percentage under $H_a : \beta_1 = 0.9,\ \rho = 0.$

T	$Test$	Size 10%	5%	1%
2	$DISJ$	80.25	52.60	11.50
	$OVER$	83.80	53.30	11.20
	$WHITE$	100.00	100.00	100.00
10	$DISJ$	100.00	100.00	100.00
	$OVER$	100.00	100.00	100.00
	$WHITE$	100.00	100.00	100.00
50	$DISJ$	100.00	100.00	100.00
	$OVER$	100.00	100.00	100.00
	$WHITE$	100.00	100.00	100.00

Conclusions and Discussions

• We propose a Wald test for the parameter in a linear regression model, in which there are cross-sectional correlations among the N units, where$N$ goes to infinity.

• Unlike the existing literature,
(i) We do not assume away the cross-sectional correlations.
(ii) We do not assume the number of time-series units, denoted as T is large, as long as $T \geq 2$.
(iii) We do not rely on the definition of economic distance.
(iv) Our approach is applicable to a general linear regression model.

• In one of the sections, we also consider a unit root test and a test for cointegration.

- In future research:

(i) We will consider the case where, *possibly*, $T \to \infty$.

(ii) The optimal choice of $\#\mathcal{T}_1$ and/or $\#\mathcal{T}_2$.