Banded Spatio-Temporal Autoregressions

Litian Xu	Hairong Hu	Xuan Ai
PB18081662	PB18051089	PB18010392

Contents

Abstract

We employ a new class of spatio-temporal models with unknown and banded autoregressive coefficient matrices,which is proposed by Gao et al.(2018).The setting represents a sparse structure for high-dimensional spatial panel dynamic models when panel members represent economic individuals at many different locations.The proposal is radically different from the existing literature on the inference for high-dimensional banded covariance matrices since the implied autocovariance matrices are unlikely to be banded.We apply the least squares method based on a Yule-Walker equation to estimate autoregressive coefficient matrices.The estimators based on multiple Yule-Walker equations are also studied.We also proposed a ratio-based method for determining the bandwidth of autoregressive matrices.The proposed methodology is further illustrated using simulated data sets.

Keywords: Banded coefficient matrices,Least squares estimation,Spatial panel dynamic models,Yule-Walker equation.

1 Introduction

One common feature in most literature on spatial econometrics is to specify each autoregressive coefficient matrix in a spatial autoregressive or a spatial dynamic panel model as a product of an unknown scalar parameter and a known spatial weight matrix, and the focus of the inference is on those a few unknown scalar parameters placed in front of spatial weight matrices. Using spatial weight matrices reflects the initial thinking that spatial dependence measures should take into account both spatial locations and feature variables at locations simultaneously.

By drawing energy and inspiration from the recent development in sparse high-dimensional (auto)regressions (Guo et al., 2016), we propose in this paper a new class of spatio-temporal models in which autoregressive coefficient matrices are completely unknown but are assumed to be banded, i.e. the non-zero coefficients only occur within the narrow band around the main diagonals. This avoids the difficulties in specifying spatial weight matrices subjectively.The setting specifies autoregressions over neighboring locations only. It is worth pointing out that the implied autocovariance matrices are unlikely to be banded in spite of the banded autoregressive coefficient matrices.

Due to the endogeneity in spatial autoregressive models,we adapt a version of generalized method of moments estimation based on a Yule–Walker equation (Dou et al., 2016). Furthermore the estimation of the parameters based on multiple Yule–Walker equations is also investigated. The asymptotic property of the estimation is established when the dimensionality p (i.e. the number of panels) diverges together with the sample size n (i.e. the length of the observed time series). The convergence rates of the estimators are the same with those in Dou et al.(2016). More precisely, the estimated coefficients are asymptotically normal when p = $o(\sqrt{(}n))$, and is consistent when$p = o(n)$.

In practice, the width of the nonzero coefficient bands in the coefficient matrices needs to be estimated. We propose a ratio-based estimation method which is shown to lead to a consistent estimated width when both n and p tend to infinity.

This article is organized as follows: Section 2 introduces the models and associate methods.The numerical illustration are reported in Section 3.Section 4 discusses something more and summarizes.All code are relegated into an Appendix.

2 Model and estimation method

2.1 Spatio-temporal regression model

Consider the spatio-temporal regression

$$y_t = Ay_t + By_{t-1} + \epsilon_t \tag{1}$$

where $y_t = (y_{1,t}, \cdots, y_{p,t})^T$ represents the observations collected from p locations at time t,$\epsilon_t = (\epsilon_{1,t}, \cdots, \epsilon_{p,t})$ is the innovation at time t and satisfies the condition that

$$E(\epsilon_t) = 0, Var(\epsilon_t) = \Sigma_\epsilon, cov(y_{t-j}, \epsilon_t) = 0, \forall j \geq 1$$

where Σ_ϵ is an unknown positive definite matrix.We assume that A and B are $p * p$ unknown banded coefficient matrices,i.e.,

$$a_{i,j} = b_{i,j} = 0, \forall |i-j| > k_0 \tag{2}$$

and $a_{i,j} = 0, \forall 1 \le i \le p$.In the above model(1),A captures the pure spatial dependency among different locations,and B captures the dynamic dependency.

Note that the conditionn $a_{i,j} = b_{i,j} = 0$ does not imply $Cov(y_{i,t}, y_{j,t}) = 0$ or $Cov(y_{i,t}, y_{j,t-1}) = 0$.

Let $I_p - A$ be invertible,and all the eigenvalues of $(I_p - A)^{-1}B$ be smaller that 1 in modulus.Then model(1) can be rewritten as

$$y_t = (I_p - A)^{-1} B y_{t-1} + (I_p - A)^{-1} \epsilon_t \tag{3}$$

the Yule-Walker equations are

$$\Sigma_0 = (I_p - A)^{-1} B \Sigma_1^T + (I_p - A)^{-1} \Sigma_\epsilon (I_p - A^T)^{-1}, \Sigma_j = (I_p - A)^{-1} B \Sigma_{j-1}, for \quad j \ge 1 \tag{4}$$

where $\Sigma_j = (y_{t+j}, y_t)$ for all $j \ge 0$.

In this paper,y_t is a stationary process defined by (3).

2.2 Generalized Yule-Walker estimation

The Yule-Walker equation imlpies

$$\Sigma_1^T e_i = \Sigma_1^T a_i + \Sigma_0 b_i = V_i \beta_i, \quad i = 1, \cdots, p, \tag{5}$$

where e_i denotes the $p * 1$ unit vector with 1 as its i-th element,$_i$ is the $\tau_i * 1$vector obtained by stacking together the non-zero elements in a_i and b_j,and V_i is the $p * \tau_i$ matrix consisting of the corresponding columns of Σ_1^T and Σ_0.

$$\tau_i(k_0) = \begin{cases} 2(k_0 + i) - 1, & 1 \leq i_0 \\ 4k_0 + 1, & k_0 < i \leq p - k_0 \\ 2(k_0 + p - i) + 1, & p - k_0 < i \leq p \end{cases}$$

We first treat the bandwidth k_0 as a known parameter and apply a version of generalized method of moment estimated based on (5).We apply least squares method to estimate (A,B) by solving the following minimization problems.

Let

$$\bar{\Sigma}_1 = \frac{1}{n}\Sigma_{t=2}^{n} y_t y_{t-1}^T \quad and \bar{\Sigma}_0 = \frac{1}{n}\Sigma_{t=2}^{n} y_{t-1} y_{t-1}^T \tag{6}$$

let $\bar{z}_i = \bar{\Sigma}_i^T e_i$ and $\bar{V}_i$ be the sample version of V_i,leads to the least square estimator

$$\bar{\beta}_i = (\bar{V}_i^T \bar{V}_i)^{-1} \bar{V}_i^T \bar{z}_i, \quad i = 1, \cdots, p. \tag{7}$$

The corresponding residual sum of squares is

$$RSS_i(k_0) = \frac{1}{p} ||\bar{z}_i - \bar{V}_i \bar{\beta}_i||^2, \quad i = 1, \cdots, p. \tag{8}$$

2.3 A root-n consistent estimator for large p

When p is large,the number of parameters to be estimated is far less than the number of estimation equations, which is so-called an over-determined case. To solve the problem, we need reduces the number of the estimation equations. The following alternative estimator we propose can reduce it from p to a smaller constant.

Note that ,the l-th row of $\bar{V}_i$ is $e_l^T \bar{V}_i$.According to the definition of V_i, $\bar{V}_i$ is the sample covariance between $y_{l,t-1}$ and $u_{t,i}$ and it can be expressed as $\frac{1}{n} e_l^T \sum_{t=2}^{n} y_{t-1} u_{t,i}^T$, where $u_{t,i}$ is a $\tau_i \times 1$ vector consisting of $y_{j,t}$ for $j \in \{j :$

$1 \leq j \leq p, 1 \leq |j-i| \leq k_0\} := S_i$ and $y_{l,t-1}$ for $l \in \{j : 1 \leq j \leq p, |j-i| \leq k_0\} := S_i^+$. For convenience, we mark the length of $u_{t,i}$ as τ in our code, and $\tau = \tau 1$(number of $y_{j,t}$ in $u_{t,i}$) + $\tau 2$(number of $y_{l,t-1}$ in $u_{t,i}$)

Then, to solve the above over-determined case, we measure the strength of the correlation between $y_{l,t-1}$ and $u_{t,i}$ by

$$\delta_l^{(i)} = \frac{1}{n}\Sigma_{t=2}^n(\Sigma_{j\in S_i}|y_{l,t-1}y_{j,t}| + \Sigma_{j\in S_i^+}|y_{l,t-1}y_{j,t-1}|) \tag{9}$$

We only keep the d_i elements in y_t with largest $\delta_l^{(i)}$ in the l-th equation, because when $\delta_l^{(i)}$ is close to 0, it carry little information on β_i. We mark the sub-vector of y_t as w_t^i (w_t^i is a $d_i \times 1$ vector).

Similar to the composition of $\hat{V}_i$ and $\hat{z}_i$, we let $\tilde{W}_i = \frac{1}{n}\Sigma_{t=2}^n w_{t-1}^i u_{t,i}^T$ and $\tilde{z}_i = \frac{1}{n}\Sigma_{t=2}^n w_{t-1}^i y_{i,t}$. Then we obtain the least square estimation of β_i

$$\tilde{\beta}_i = (\tilde{W}_i^T\tilde{W}_i)^{-1}\tilde{W}_i^T\tilde{z}_i, i = 1, ..., p \tag{10}$$

2.4 Bandwidth parameter k0

In practice, the bandwidth parameter k0 is unknown. We propose below a method to estimate it.

If we replace $(\hat{\Sigma}_0, \hat{\Sigma}_1)$ by the true (Σ_0, Σ_1), the corresponding true value of $RSS_i(k)$ is positive and finite for $1 \leq k \leq k0$,and is equal to 0 for $k_0 \leq k \leq K$. Thus the ratio $RSS_i(k1)/RSS_i(k)$ is finite for $k < k_0$,$RSS_i(k_0 1)/RSS_i(k_0)$ is excessively large, and $RSS_i(k1)/RSS_i(k)$ is effectively '0/0' for $k > k_0$.To avoid the singularities when $k > k_0$, we introduce a small factor $w_n = C/n$ in the ratio for some constant C > 0. A ratio-based estimator for k_0 is defined as

$$\hat{k} = max_{1\leq i\leq p} \arg\max_{1\leq k\leq K} \frac{RSS_i(k-1) + w_n}{RSS_i(k) + w_n} \tag{11}$$

2.5 Estimation with multiple Yule-Walker equation

We propose an alternative estimation for the over-determined case in section 2.3, but it's still not more accurate since we only make use of partial information for the parameters. Only one Yule-Walker equation is used to estimate parameter in previous sections, we consider to use multiple Yule-Walker equations in this section.

Let r be a prescribed positive integer. We can obtain more information in following r Yule-Walker equations. According to $(4): \Sigma_j^T = \frac{1}{n}\Sigma_{t=j+1}^n y_t y_{t-j}^T, j = 1, ..., p$

$$\begin{bmatrix} \Sigma_1^T \\ \Sigma_2^T \\ ... \\ \Sigma_r^T \end{bmatrix} = \begin{bmatrix} \Sigma_1^T \\ \Sigma_2^T \\ ... \\ \Sigma_r^T \end{bmatrix} A^T + \begin{bmatrix} \Sigma_0^T \\ \Sigma_1^T \\ ... \\ \Sigma_{r-1}^T \end{bmatrix} B^T \tag{12}$$

Denote

$$\hat{x}_i = \begin{bmatrix} \hat{\Sigma}_1^T \\ \hat{\Sigma}_2^T \\ ... \\ \hat{\Sigma}_r^T \end{bmatrix} \tag{13}$$

and

$$\hat{G} = \begin{bmatrix} \hat{\Sigma}_1^T & \hat{\Sigma}_0^T \\ \hat{\Sigma}_2^T & \hat{\Sigma}_1^T - \frac{1}{n} y_{n-1} y_n^T \\ ... & ... \\ \hat{\Sigma}_r^T & \hat{\Sigma}_{r-1}^T - \frac{1}{n} y_{n-r+1} y_n^T \end{bmatrix} \tag{14}$$

The problem need to be solved is still the minimization formula

$$\underset{\theta_i}{min} \, ||\hat{x}_i - \hat{G}_i \theta_i||_2^2 \tag{15}$$

We apply the least square method to estimate (A,B),where θ_i is a $\tau_i \times 1$ vector and $\hat{G}_i$ is the $rp \times \tau_i$ submatrix of $\hat{G}$ corresponding to the nonzero elements of a_i and b_i. Let $\hat{\hat{\beta}}_i, \hat{\hat{A}}, \hat{\hat{B}}$ be the estimation of β_i, A, B

$$\Rightarrow \qquad \hat{\hat{\beta}}_i = (\hat{G}_i^T \hat{G}_i)^{-1} \hat{G}_i^T \hat{x}_i, \qquad i = 1, ..., p \tag{16}$$

3 Simulation

We simulate y_t from model with independent and N(0,1) innovations $\epsilon_{i,t}$.

For each model, we set sample size n = 500, 1000, and 2000 and dimension of time series p = 100, 300, 500, 800 and 1000. This leads to the 15 different (n, p) combinations. For each setting, we replicate the experiment 100 times, and calculate the relative frequencies for the occurrence of events $k = k_0$, $k > k_0$ and $k < k_0$ in the 500 replications. We also calculate the means and the standard deviations of the estimation errors $||A\hat{A}||_2$ and $||B - \hat{B}||_2$.

3.1 case1

3.1.1 settings

Elements $a_{i,j}, b_{i,j}$ for $|i - j| = k_0$ are drawn independently from uniform distribution on two points $\{-2, 2\}$, and $a_{i,j}$ for $0 < |i - j| < k_0$ and $b_{i,j}$ for $|i - j| < k_0$ are drawn independently from the mixture distribution $wI_0 + (1w)N(0, 1)$ with $P(w = 1) = 0.4 = 1 - P(w = 0)$. We then rescale A and B to $n1 * ||A - \hat{A}||_2$ and $n2 * ||B - \hat{B}||_2$, where n1 and n2 are drawn independently from $U[0.4, 0.8]$.

3.1.2 results

Relative frequencies of the occurrence of the events $k = k_0$, $k < k_0$, and $k > k_0$ based on r=1 for Case1 with $k_0 = 3, K = 10, r = 1$ are showed in Table1(in the end of the paper).

Mean and standard deviations(in parentheses) of $||A - \hat{A}||_2$ and $||B - \hat{B}||_2$ for Case1 with $k_0 = 3, K = 10$ and $r = 1, 2, 3$ are showed in Table2(in the end of the paper).

Table 1: Relative frequencies of the occurrence of the events $k = k_0$, $k < k_0$, and $k > k_0$ based on r=1 for Case1 with k0=3,K=10

			r=1		
p	n	SNR	$k = k_0$	$k > k_0$	$k < k_0$
50	500	1,178	0.00	1.00	0.00
50	1000	1.169	0.01	0.89	0.10
50	2000	1.179	0.02	0.81	0.17
100	500	1.153	0.03	0.93	0.04
100	1000	1.152	0.00	0.84	0.16
100	2000	1.146	0.03	0.77	0.20

3.1.3 analysis

As indicated in the table1,when the sample size n increases,the errors in estimating the coefficient matrices A and B decreases while the relative frequency for the correct specification of the bandwidth parameter k_0 increase. Note that the errors in estimating A and B based on r=1,2,3 show no clear difference.Due to the limitation of computation ability of personal computers,we just simulate two different dimensions.There is no clear pattern in performance with respect to different values of the dimension p.This is due to the fact that the signal-to-noise ratio does not vary monotonically with respect to p.

3.2 case2

3.2.1 settings

Elements $a_{i,j}$, $b_{i,j}$ for $|i-j| = k_0$ are drawn independently from $U([2.5, 1.5] \cup [1.5, 2.5])$, and $a_{i,j}$ for $0 < |i-j| < k_0$ and $b_{i,j}$ for $|i-j| < k_0$ are drawn independently from $U[1, 1]$. We then rescale A and B as in Case 1 above.

3.2.2 results

Relative frequencies of the occurrence of the events $k = k_0$, $k < k_0$, and $k > k_0$ based on r=1 for Case2 with $k_0 = 3, K = 10, r = 1$ are showed in Table3.

Mean and standard deviations(in parentheses) of $||A-\hat{A}||_2$ and $||B-\hat{B}||_2$ for Case2 with $k_0 = 3, K = 10$ and $r = 1, 2, 3$ are showed in Table4.

Table3 and Table 4 are showed in the end of paper.

3.2.3 analysis

As indicated in the table2,when the sample size n increases,the errors in estimating the coefficient matrices A and B decreases while the relative frequency for the correct specification of the bandwidth parameter k_0 increase.Note that the errors in estimating B based on r=1,2,3 show no clear difference.However,when n and p are fixed,the errors in estimating A decreases with r,which is inverse conclusions against the paper.Due to the limitation of computation ability of personal computers,we just simulate two different dimensions.There is no clear pattern in performance with respect to different values of the dimension p.This is due to the fact that the signal-to-noise ratio does not vary monotonically with respect to p.

3.3 Comparison between two estimators

To compare the estimators in Section 2.2 and Section 2.3.we generate the data as Case2 with $K = 5$ and $k_0 = 1$.We denote the two estimators by Estimator 1 and Estimator 2 respectively.We can see that for each p,the estimation errors decrease as the sample size increases.On the other hand,for each n and p,the root-n consistent estimator(Estimator 2) tends to have larger estimation errors since only makes use of part of the information for the parameters as long as $d_i < p$.

3.3.1 setting

We generate the data as Case 2 with K = 5 and k0 = 1. For each p = 50, 75 and 100, we set the sample size n = 500,1000 and 2000 respectively. In addition, we choose $d_i = min(p; [n^{0:495}])$ and denote the two estimators by Estimate I and Estimator II, respectively.

3.3.2 results

Relative frequencies of the occurrence of the events $k = k_0$, $k < k_0$,and $k > k_0$ based on r=1,and mean and standard deviations(in parentheses) of $||A - \hat{A}||_2$ and $||B - \hat{B}||_2$ for Case2 with $k_0 = 1, K = 5, r = 1$ and $d_i = min(p, [n^{0.495}])$ are showed in Table5.

(For the limit of space,we show Table5 in the end of the paper)

4 Real data

Now we consider the annual mortality rates in the period of 1872-2009 of the Italian population at age i, for i = 10,11,...,50. The data is from http://www.mortality.org/.

Let $m_{i,t}$ be the original mortality rate at age i in the t-th year. Fig.1 displays the time series of $m_{i,t}$ with age i = 10,30 and 50 respectively. Let $y_{i,t}$,t = 1872,...,2009 be the centered log-scaled mortality rates for the i-th age group, i = 10,11,...,50, $p = 41$ and $n = 138$. The order of $y_{i,t}$ is natural.

We use the ratio-based method to estimate bandwidth parameter k for this data set, and compute one-step ahead predictive errors for the last 9 data points for each of 41 series. The estimated error is 0.00896, indicating clearly that the proposed banded model performs well.

5 Summary

5.1 Difficulties encountered and solutions

1) There's few difficulties in comprehending how to build the Spatio-Temporal model and obtain the Yule-Walker equations after learning the Analysis of Time Series this semester at first. But it's a little hard

to understand several theoretical properties. Some processing of data is a bit confusing such as removing the last term of $\hat{\Sigma}_j^T$ in the second half columns of $\hat{G}$ in the equation (14) in section 2.5.

2) More problems occurred when editing code for the spatio-temporal model. We enumerate some of them as follows.

   * When computing the least square estimation of β_i, the inverse of $V_i^T V_i$ or $W_i^T W_i$ often shows singular, which leads to be insoluble. We transfer to solve the generalized inverse of matrix in the end.

   * When generating data in case1 and case2, at first we respectively call function to generate Y, A and B, which greatly increase the time spent, decreasing efficiency and even bring about some errors. Afterwards we assign the result of ”generate-data function” to ”S”, then take what we want from ”S”. The solution to use the estimation from ”largep function” is similar.

   * In section 3.3, we need to compare two estimators respectively using the method in section 2.22.3. The signal-to-noise ratio(SNR),the norm of $A - \hat{A}$ and $B - \hat{B}$ and their standard error of estimation are mean value of 500 times repeated experiments. However, it’s time-consuming and likely to breakdown of our code and computer especially for large sample situation. According to the paper, we can properly reduce the total times of cycle to 100 times.

   * Due to sample capacity and high dimension, it’s hard to clarify the complicated subscripts in numerous expressions and put it into practice when editing code,let alone discovering the minor error in many lines of code. One false step will make a great difference.

5.2 Gains and Reflections

It's an exciting experience to put what we learned in the class into practice. Banded Spatio-Temporal model is a good example to use Yule-Walker function. We need more patience and curiosity to explore various models and seek solutions, learn to think questions deeply from many perspectives so that we can obtain more suitable model and estimation.

5.3 Concluding remarks

We propose in this paper a new class of banded spatio-temporal models. The setting does not require pre-specified spatial weight matrices. The coefficient matrices are estimated by a generalized method of moments estimation based on a Yule-Walker equation. The bandwidth of the coefficient matrices is determined by a ratio-based method.

6 Reference

References

[1] Zhaoxing Gao,Yingying Ma,Hansheng Wang,Qiwei Yao. Banded Spatio-Temporal Autoregressions. Department of Statistics, London School of Economics, United Kingdom.School of Economics and Management, Beihang University, China.Guanghua School of Management, Peking University, China,2018:4.

Table 2: mean and standard deviations(in parentheses) of $||A-\hat{A}||_2$ and $||B-\hat{B}||_2$ for Case1 with k0=3,K=10 and r=1,2,3 respectively

		r=1		**r=2**		**r=3**	
p	n	$\|\|A-\hat{A}\|\|_2$	$\|\|B-\hat{B}\|\|_2$	$\|\|A-\hat{A}\|\|_2$	$\|\|B-\hat{B}\|\|_2$	$\|\|A-\hat{A}\|\|_2$	$\|\|B-\hat{B}\|\|_2$
50	500	2.916(0.710)	1.502(0.274)	2.868(0.672)	1.496(0.278)	2.961(0.724)	1.467(0.279)
50	1000	2.590(0.754)	1.294(0.272)	2.613(0.783)	1.282(0.233)	2.668(0.778)	1.348(0.294)
50	2000	2.323(0.855)	1.201(0.273)	2.426(0.831)	1.237(0.276)	2.399(0.702)	1.239(0.283)
100	500	1.475(0.305)	1.090(0.181)	1.124(0.199)	1.054(0.129)	1.433(0.309)	1.096(0.142)
100	1000	1.299(0.289)	0.976(0.162)	1.042(0.175)	0.917(0.149)	1.329(0.251)	0.998(0.150)
100	2000	1.280(0.257)	0.897(0.158)	0.997(0.190)	0.885(0.160)	1.270(0.273)	0.922(0.169)

Table 3: Relative frequencies of the occurrence of the events $k = k_0$, $k < k_0$, and $k > k_0$ based on r=1 for Case2 with k0=3,K=10

			r=1		
p	n	SNR	$k = k_0$	$k > k_0$	$k < k_0$
50	500	1.208	0.02	0.94	0.04
50	1000	1.198	0.00	0.90	0.10
50	2000	1.204	0.00	0.82	0.18
100	500	1.202	0.012	0.852	0.136
100	1000	1.204	0.040	0.650	0.310
100	2000	1.19	0.040	0.700	0.260

Table 4: mean and standard deviations(in parentheses) of $||A - \hat{A}||_2$ and $||B - \hat{B}||_2$ for Case2 with k0=3,K=10 and r=1,2,3 respectively

		r=1		**r=2**		**r=3**	
p	n	$\|\|A - \hat{A}\|\|_2$	$\|\|B - \hat{B}\|\|_2$	$\|\|A - \hat{A}\|\|_2$	$\|\|B - \hat{B}\|\|_2$	$\|\|A - \hat{A}\|\|_2$	$\|\|B - \hat{B}\|\|_2$
50	500	1.070(0.129)	0.983(0.185)	0.937(0.123)	0.996(0.177)	0.854(0.121)	0.925(0.171)
50	1000	0.975(0.127)	0.984(0.145)	0.957(0.156)	0.987(1.173)	0.836(0.124)	0.913(0.159)
50	2000	0.960(0.165)	0.887(0.178)	0.939(0.147)	0.930(0.172)	0.876(0.169)	0.902(0.194)
100	500	0.933(0.134)	0.984(0.170)	0.824(0.130)	0.984(0.147)	0.772(0.142)	0.973(0.162)
100	1000	0.907(0.132)	0.974(0.174)	0.842(0.152)	0.974(0.180)	0.802(0.139)	0.932(0.177)
100	2000	0.938(0.155)	0.916(0.162)	0.821(0.138)	0.895(0.165)	0.813(0.127)	0.960(0.170)

Table 5: mean and standard deviations(in parentheses) of $||A - \hat{A}||_2$ and $||B - \hat{B}||_2$ for Case2 with $k0 = 1, K = 5$ and $d_i = \min(p, n^{0.495})$ respectively

			r=1			**Estimator 1**		**Estimator 2**	
p	n	SNR	$k = k_0$	$k > k_0$	$k > k_0$	$\|\|A - \hat{A}\|\|_2$	$\|\|B - \hat{B}\|\|_2$	$\|\|A - \hat{A}\|\|_2$	$\|\|B - \hat{B}\|\|_2$
50	50	1.246	0	1	0	1.021(0.202)	1.275(0.243)	9.98(12.046)	10.809(12.151)
50	100	1.247	0	1	0	0.952(0.150)	1.091(0.170)	2.378(0.679)	2.786(0.675)
75	50	1.325	0	1	0	0.999(0.159)	1.237(0.162)	10.915(6.968)	12.032(6.032)
75	100	1.223	0	1	0	0.949(0.154)	1.086(0.158)	2.533(0.680)	3.133(0.778)