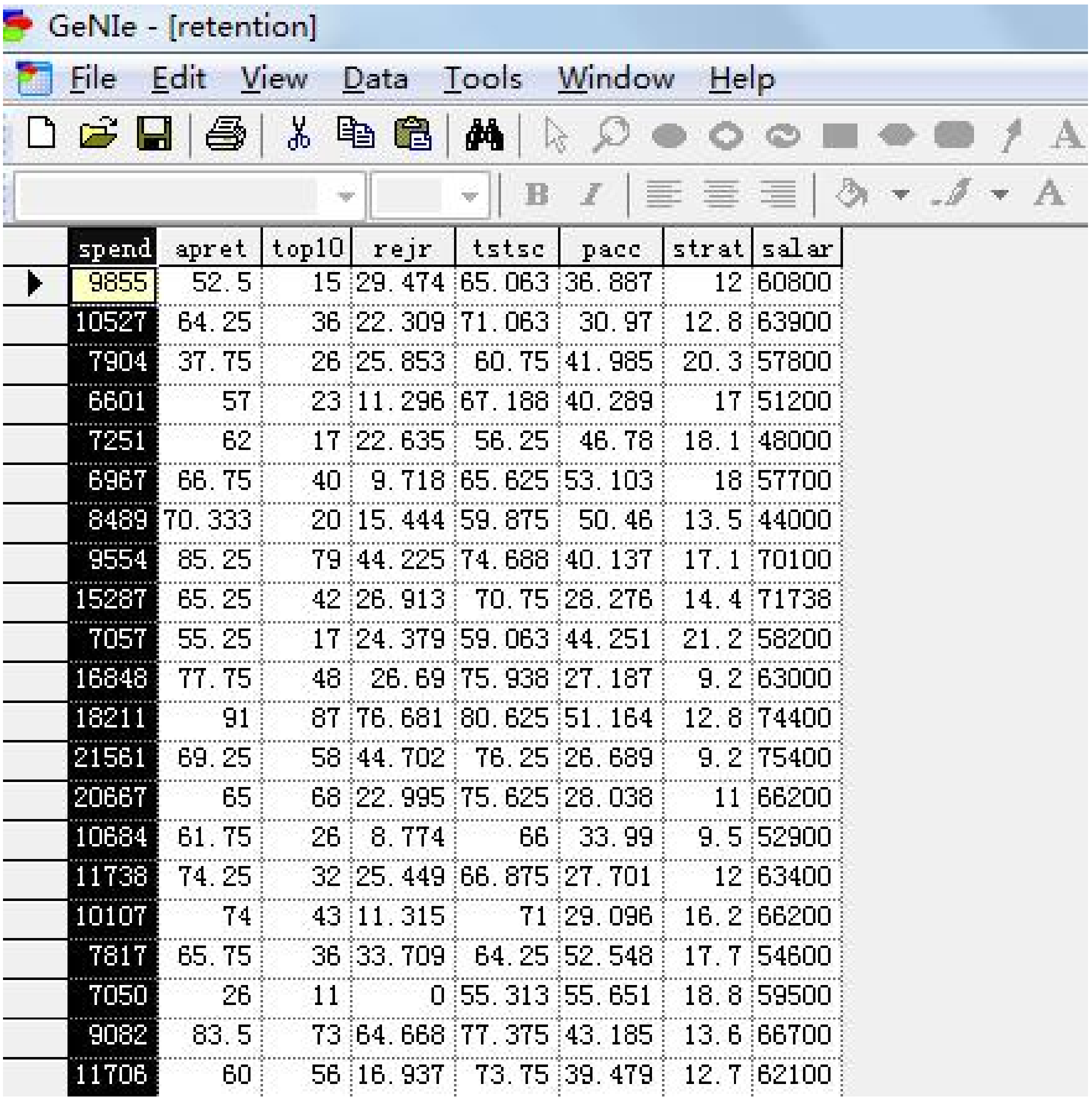
Assignment 5

Yue Su (yus55)   Dan Sun (das225)

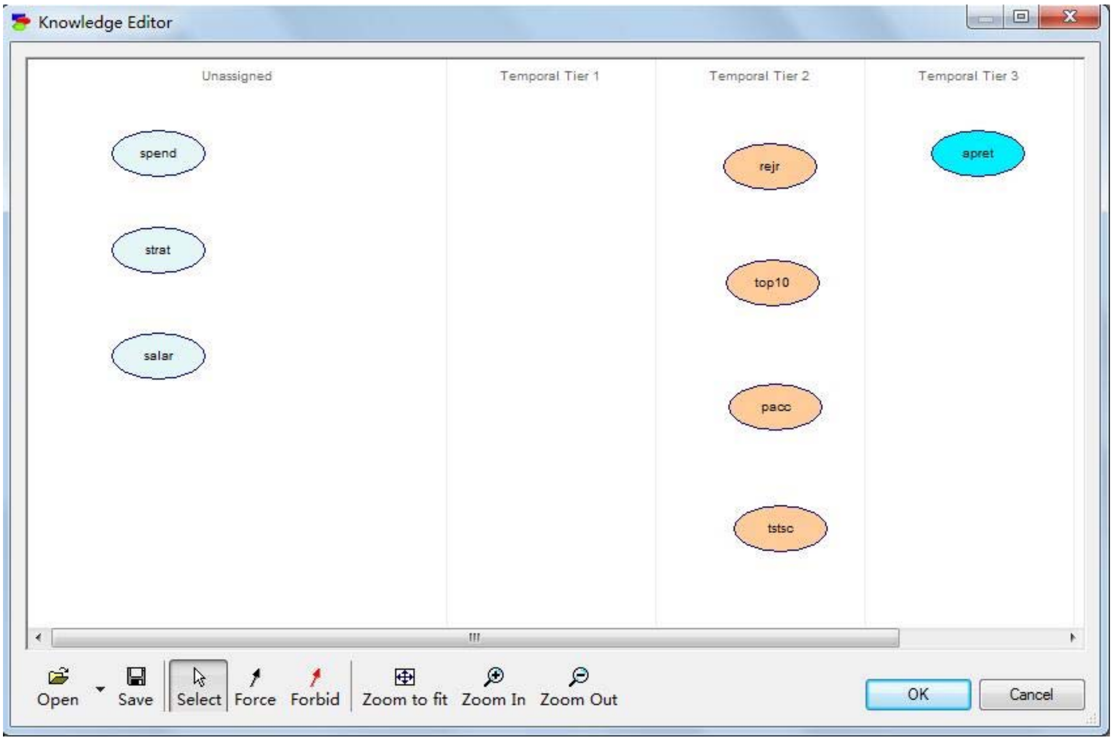
1. *Load the data*

Load the data into system. Screen shots below shows the process.



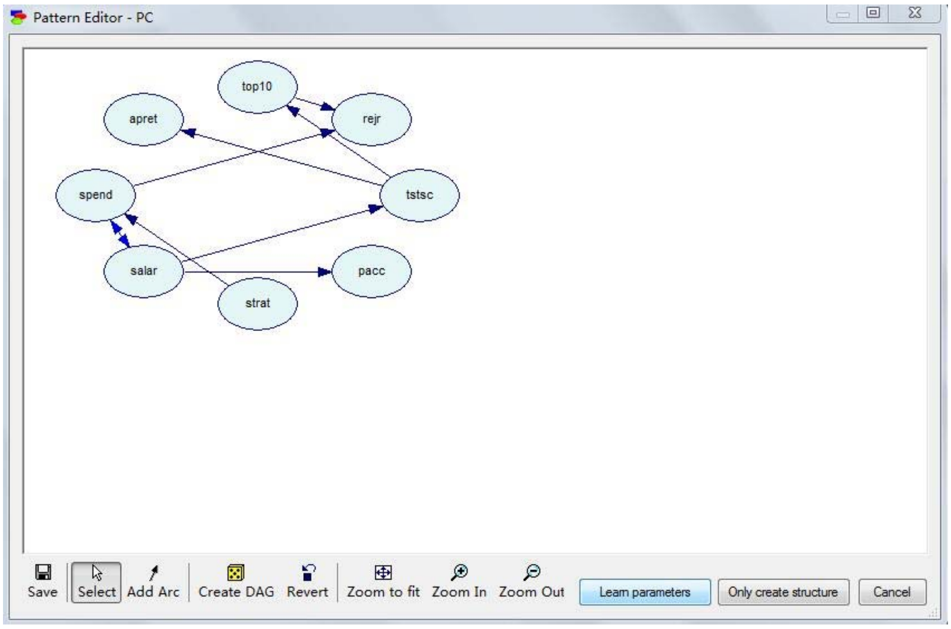
spend	apret	top10	rejr	tstsc	pacc	strat	salar
9855	52.5	15	29.474	65.063	36.887	12	60800
10527	64.25	36	22.309	71.063	30.97	12.8	63900
7904	37.75	26	25.853	60.75	41.985	20.3	57800
6601	57	23	11.296	67.188	40.289	17	51200
7251	62	17	22.635	56.25	46.78	18.1	48000
6967	66.75	40	9.718	65.625	53.103	18	57700
8489	70.333	20	15.444	59.875	50.46	13.5	44000
9554	85.25	79	44.225	74.688	40.137	17.1	70100
15287	65.25	42	26.913	70.75	28.276	14.4	71738
7057	55.25	17	24.379	59.063	44.251	21.2	58200
16848	77.75	48	26.69	75.938	27.187	9.2	63000
18211	91	87	76.681	80.625	51.164	12.8	74400
21561	69.25	58	44.702	76.25	26.689	9.2	75400
20667	65	68	22.995	75.625	28.038	11	66200
10684	61.75	26	8.774	66	33.99	9.5	52900
11738	74.25	32	25.449	66.875	27.701	12	63400
10107	74	43	11.315	71	29.096	16.2	66200
7817	65.75	36	33.709	64.25	52.548	17.7	54600
7050	26	11	0	55.313	55.651	18.8	59500
9082	83.5	73	64.668	77.375	43.185	13.6	66700
11706	60	56	16.937	73.75	39.479	12.7	62100

2. *Enter the background knowledge by clicking the Background Knowledge button. Here you can force and forbid causal connections and also order variables in temporal tiers.*

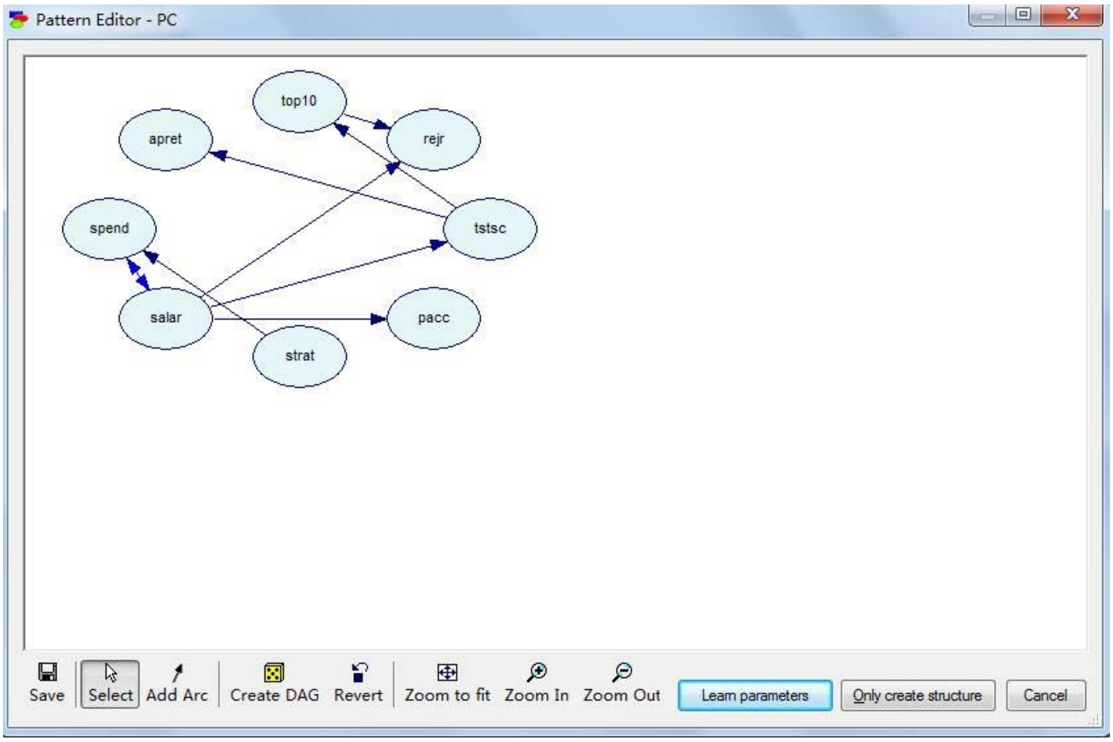


3. Choose the PC algorithm and set the significance level (PC algorithm uses classical statistical independence tests and here is where you set the significance level for these tests)

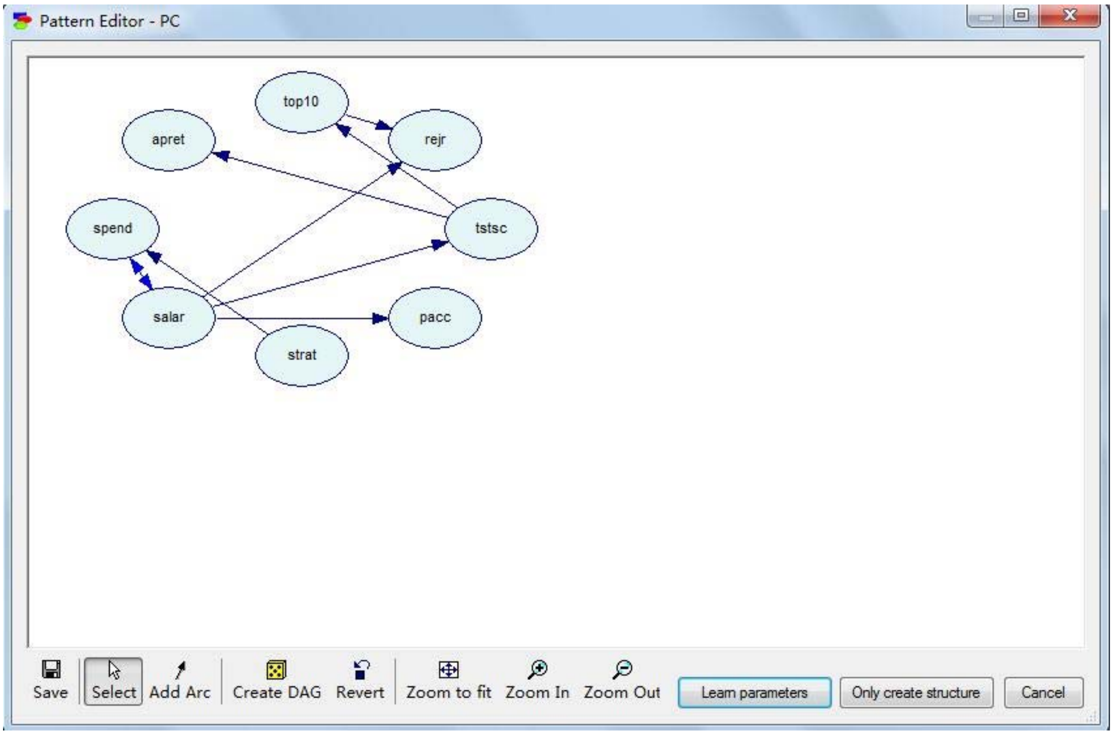
- Significance level = 0.0001



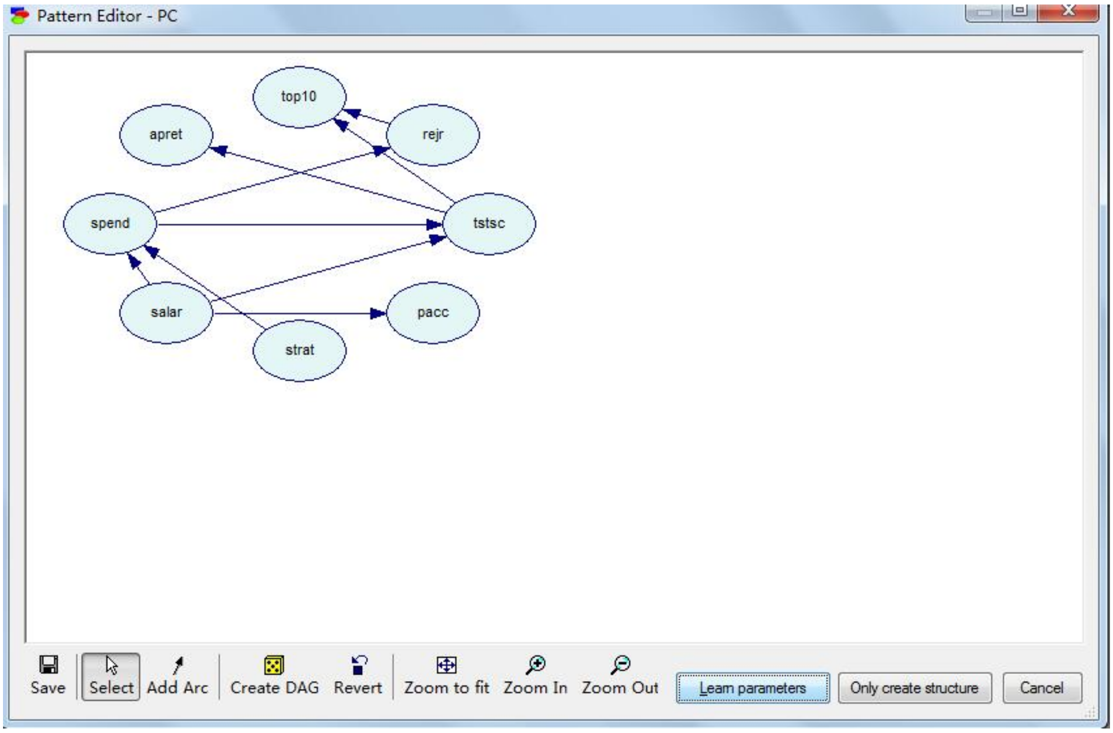
- Significance level = 0.0005



- Significance level = 0.001



- Significance level = 0.005



Depending on the significance level used in tests above, statistical decisions regarding independence may be different and a different class of causal structures may result. It is, therefore, a good practice to run the program at several significance levels. We ran algorithm with the following significance levels: p = 0.0001, 0.0005, 0.001, 0.005. The core of the structure, i.e., how freshmen retention rate and graduation rate are related to the remaining variables, was insensitive to changes in significance. This suggests that the structure proposed is robust. The edges of the graph have the following meaning: A single arrow (⟶) denotes a direct causal influence. A double headed arrow (⟷) between two variables denotes presence of a latent common cause of these two variables. An single arrow with a circle at one end (o⟶) expresses inability to deduce whether there is a direct influence between the two variables (⟶) or a latent common cause between them (⟷). An edge with circles at both ends (o—o) expresses inability to deduce whether there is a direct influence between the two variables and, if so, what is its direction or a latent common cause between them (⟷).

Most graphs contained a direct causal connection between the average test scores and freshmen retention. Also, the graphs contain latent common cause connection between spend and salary.