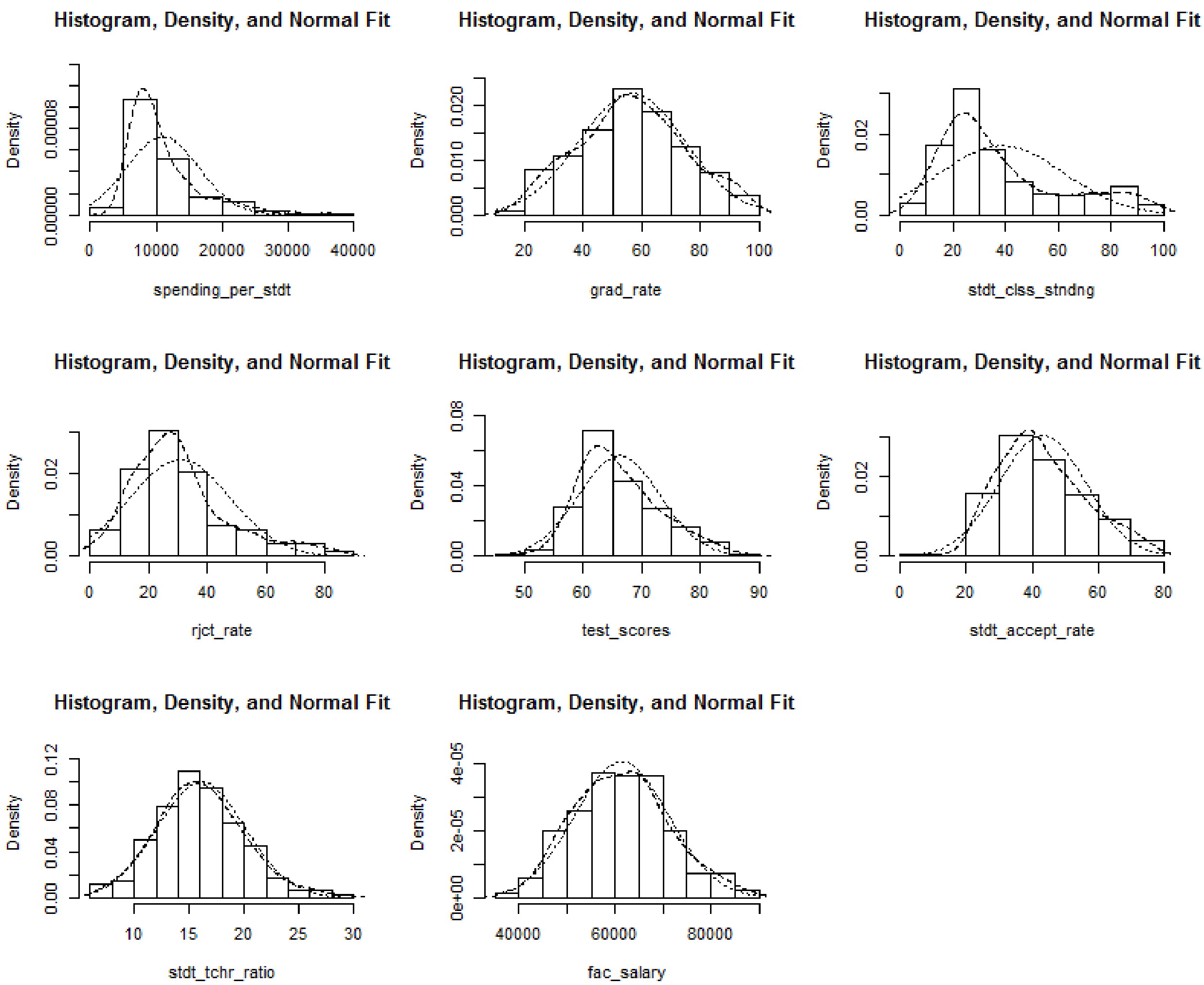
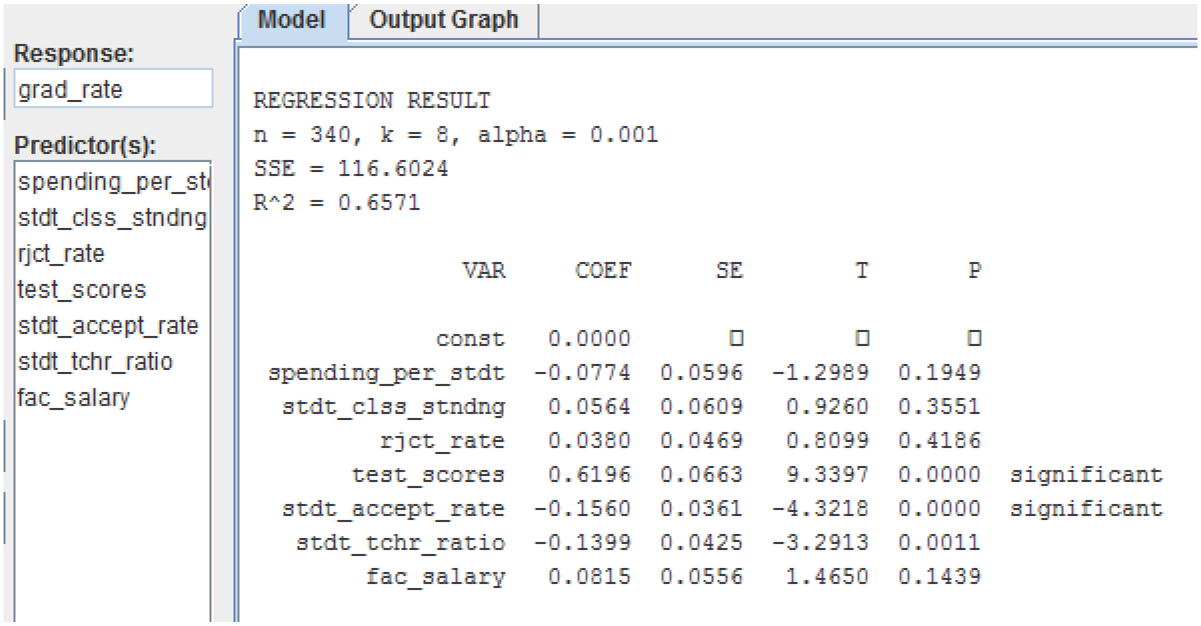
1. Assumptions: Normality



Following the above plots, we can see that it is safe to assume that the data holds on for the normality assumption as in the paper.

2. Applying Regression:

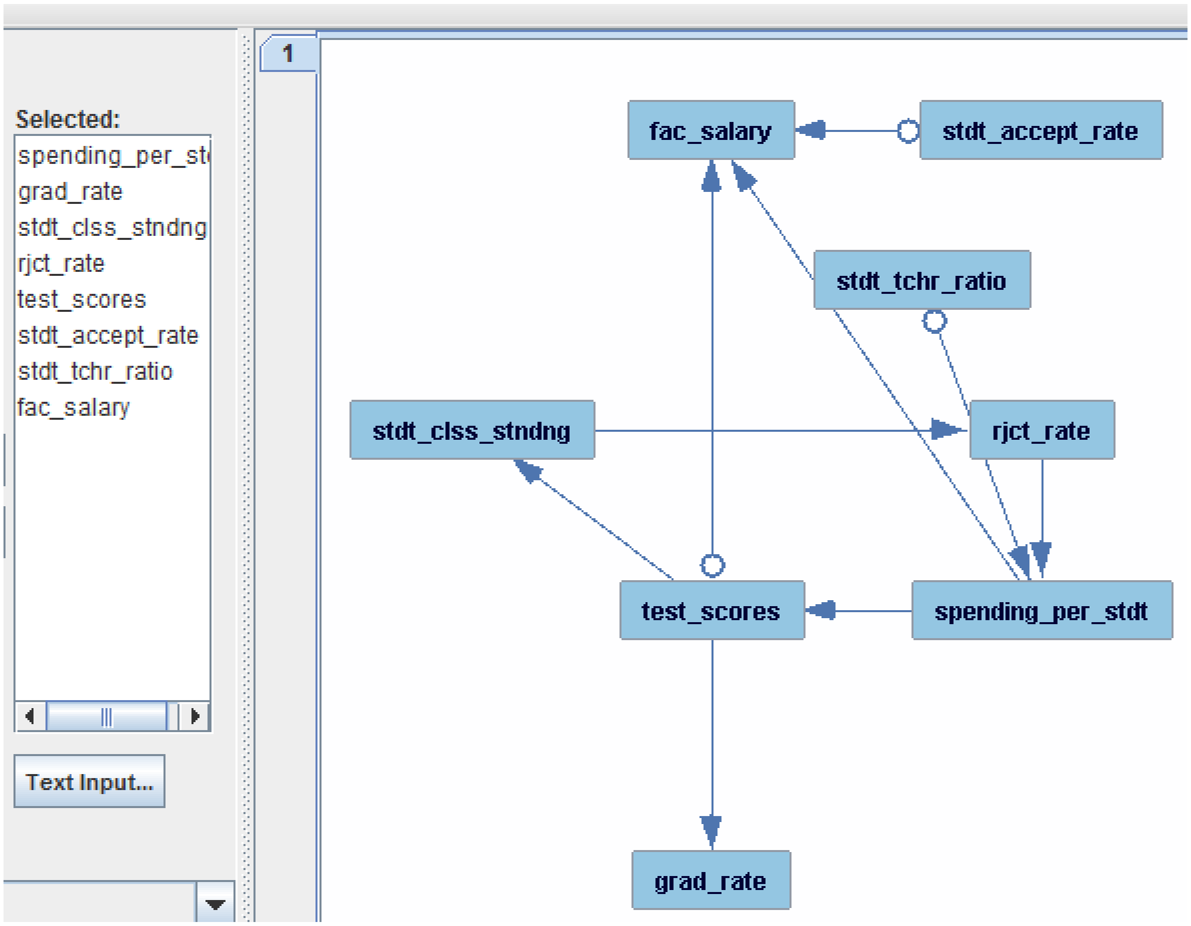


REGRESSION RESULT
n = 340, k = 8, alpha = 0.001
SSE = 116.6024
R^2 = 0.6571

VAR	COEF	SE	T	P	
const	0.0000	□	□	□	
spending_per_stdt	-0.0774	0.0596	-1.2989	0.1949	
stdt_clss_stndng	0.0564	0.0609	0.9260	0.3551	
rjct_rate	0.0380	0.0469	0.8099	0.4186	
test_scores	0.6196	0.0663	9.3397	0.0000	significant
stdt_accept_rate	-0.1560	0.0361	-4.3218	0.0000	significant
stdt_tchr_ratio	-0.1399	0.0425	-3.2913	0.0011	
fac_salary	0.0815	0.0556	1.4650	0.1439	

We can see from the result above that the regression result takes **test_scores** and **student_accept_rates** as significant in the outcome of **grad_rate.**

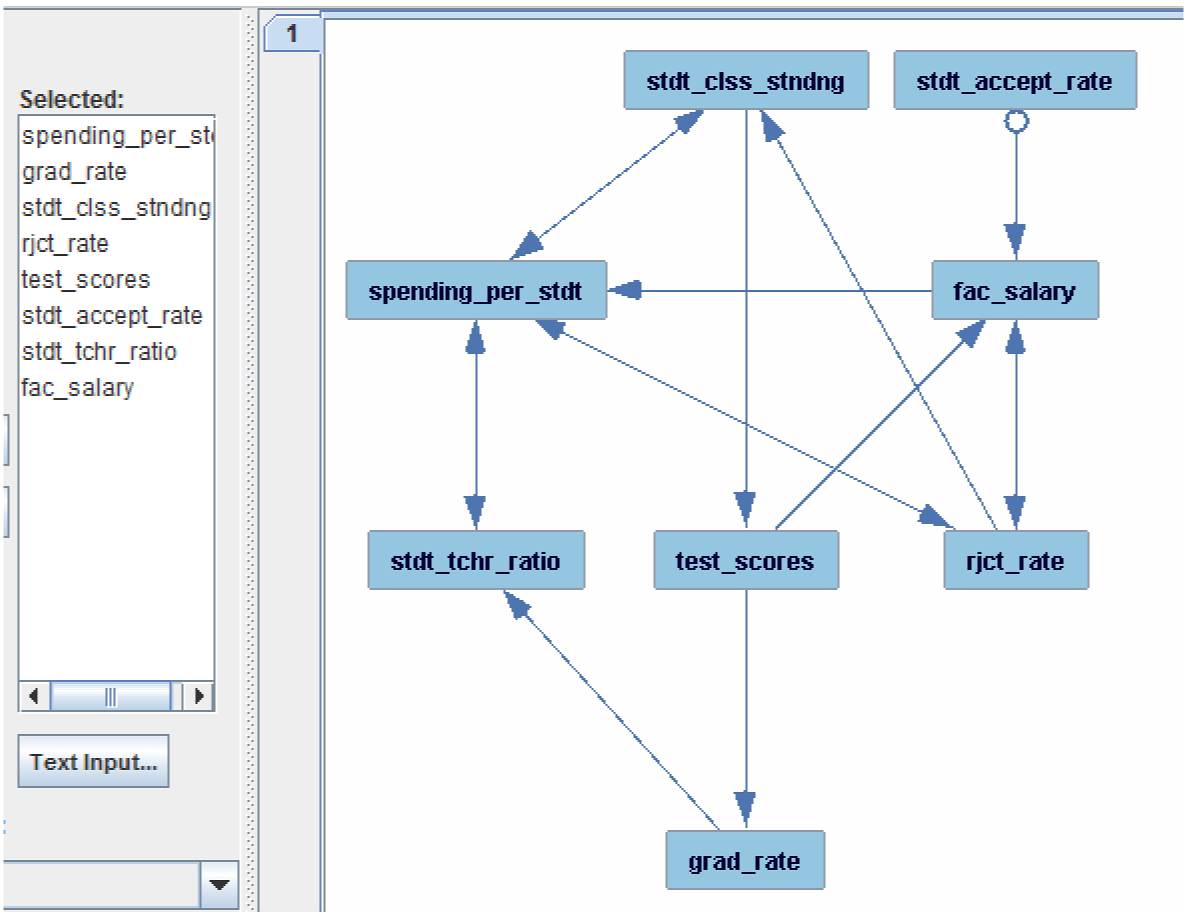
3. Now let's determine the causality using TETRAD. We start with the FCI algorithm as stated by "Drudzel" with P-value(alpha) = 0.001



As we can see, the **grad_rate** indeed bears direct causality from **test_scores** but does not directly linked to **student_accept_rate**.

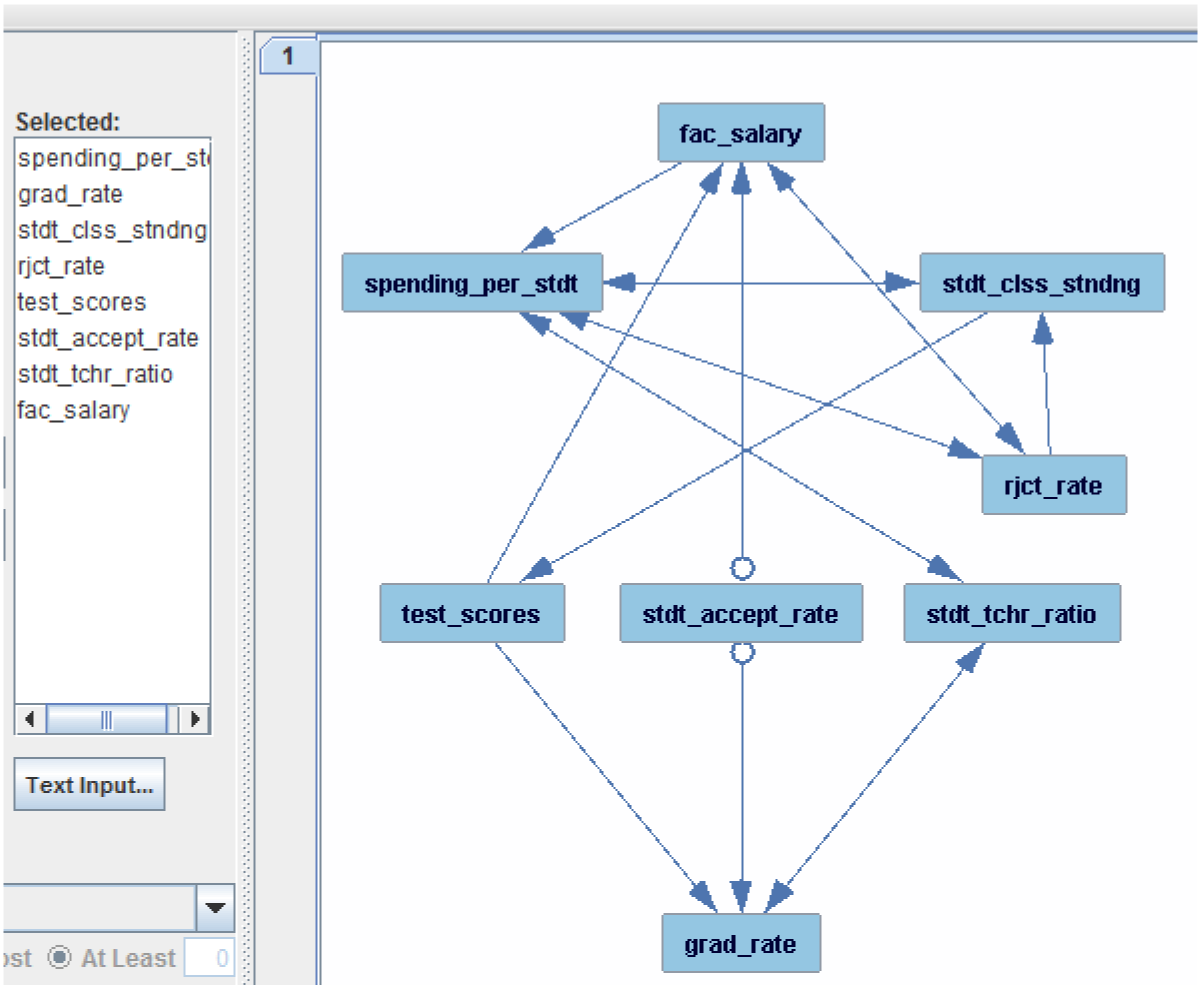
Let's change the p-value and look for any changes to the causality as suggested in the paper:

FCI using p-value = 0.05:



Increasing the p-value to 0.05 leads to a more established causality where the **grad_rate** depends only on the **test_scores.** The two-way causality to **student_teacher_ratio** as seen earlier is now tuned down and shows that **grad_rate** actually affects the **student_teacher_ratio** which makes sense and so, it is an indication of moving in the right direction.

Let's now increase the p-value to p = 0.15:



Increasing the p-value to 0.15 does infact suggest that the findings of the paper are actually correct.