Project Milestone 3

Fettah Kiran, Mohammad Imtiaz Nur

5/5/2021

Table S1:

Table S1 is based on six regression models.

The first model predicts cross topic association using the variables Yp, Zp, Kp, Wp, region count and number of discipline.

The second model predicts cross disciplinary association using the same variables except SA and replacing CIP.

The third model predicts the combination of SA and CIP.

The fourth model is similar to model-1 with additional consideration of effects publications year after 2013, and regional impact.

The model five is similar to mode-2 and model-6 is similar to model-3.

The models are created using glm library of R. All values in the table are generated by feeding the models using stargazer library.

	Dependent variable:					
	X_{SA}	X_{CIP}	$X_{SA\&CIP}$	X_{SA}	X_{CIP}	$X_{SA\&CIP}$
	(1)	(2)	(3)	(4)	(5)	(6)
y	1.032***	1.009***	1.046***	1.033***	1.021***	1.031***
	(0.0003)	(0.001)	(0.001)	(0.0004)	(0.001)	(0.001)
ε_j	0.997***	1.282***	1.415***	0.978***	1.223***	1.257***
	(0.006)	(0.008)	(0.012)	(0.006)	(0.009)	(0.009)
ln k	0.885***	1.753***	1.562***	0.897***	1.821***	1.618***
	(0.006)	(0.008)	(0.011)	(0.006)	(0.008)	(0.008)
ln w	4.655***	0.933***	4.858***	4.678***	0.929***	1.620***
	(0.009)	(0.012)	(0.016)	(0.009)	(0.012)	(0.012)
N_R	1.324***	7.810***	12.004***	1.211***	3.028***	2.782***
	(0.009)	(0.008)	(0.013)	(0.021)	(0.019)	(0.019)
N_{CIP}	1.307***			1.294***		
	(0.009)			(0.009)		
N_{SA}		1.216***			1.206***	
		(0.004)			(0.005)	
I_{2014+}				0.949***	0.754***	0.738***
				(0.020)	(0.016)	(0.017)
$I_{R_{NA}}$				0.913***	0.380***	0.383***
				(0.024)	(0.022)	(0.023)
$I_{R_{EU}}$				0.942***	0.313***	0.325***
				(0.024)	(0.023)	(0.024)
$I_{R_{AA}}$				0.746***	0.229***	0.213***
				(0.025)	(0.026)	(0.028)
I_R						
$I_{R_{NA}} \times I_{2014+}$				1.074***	1.007***	1.010***
				(0.024)	(0.023)	(0.024)
$I_{R_{EU}} \times I_{2014+}$				0.955***	1.038***	1.025***
				(0.024)	(0.024)	(0.026)
$I_{R_{AA}} \times I_{2014+}$				1.111***	0.898***	0.936***
				(0.026)	(0.034)	(0.038)
N	602,599	602,599	207,281	602,599	602,599	602,599

Exponentiated coefficients
Standard errors in parentheses
*p<0.1; **p<0.05; ***p<0.01

Table S2:

Table S2 is build on six regression models like Table S1, except now we considering neighboring topics or disciplines[CIP {1,3} with CIP {3,4-7}; SA {1} with SA {2,3, 4}] instead of considering all as previous.

The first model predicts cross topic association using the variables Yp, Zp, Kp, Wp, region count and number of discipline.

The second model predicts cross disciplinary association using the same variables except SA and replacing CIP.

The third model predicts the combination of SA and CIP.

The fourth model is similar to model-1 with additional consideration of effects publications year after 2013, and regional impact.

The model five is similar to mode-2 and model-6 is similar to model-3.

The models are created using glm library of R. All values in the table are generated by feeding the models using stargazer library.

	Dependent variable:					
	X_{SA}	X_{CIP}	$X_{SA\&CIP}$	X_{SA}	X_{CIP}	$X_{SA\&CIP}$
	(1)	(2)	(3)	(4)	(5)	(6)
y	1.030***	1.002***	1.025***	1.028***	1.012***	1.036***
	(0.0004)	(0.001)	(0.001)	(0.0005)	(0.001)	(0.001)
z_j	1.488***	1.344***	1.765***	1.428***	1.266***	1.646***
	(0.006)	(0.010)	(0.015)	(0.006)	(0.010)	(0.015)
ln k	0.530***	1.755***	1.132***	0.543***	1.832***	1.188***
	(0.006)	(0.009)	(0.013)	(0.006)	(0.009)	(0.013)
ln w	1.756***	0.889***	1.816***	1.788***	0.889***	1.799***
	(0.008)	(0.015)	(0.021)	(0.008)	(0.015)	(0.021)
N_R	1.763***	6.297***	8.424***	1.853***	2.617***	4.071***
	(0.008)	(0.009)	(0.013)	(0.017)	(0.020)	(0.028)
N_{CIP}	1.429***			1.415***		
	(0.007)			(0.007)		
N_{SA}		1.230***			1.215***	
		(0.005)			(0.005)	
I_{2014+}				1.029***	0.770***	0.795***
				(0.016)	(0.018)	(0.026)
$I_{R_{NA}}$				1.122***	0.405***	0.511***
				(0.020)	(0.025)	(0.035)
$I_{R_{EU}}$				1.192***	0.327***	0.404***
				(0.020)	(0.026)	(0.037)
$I_{R_{AA}}$				0.626***	0.172***	0.156***
				(0.022)	(0.033)	(0.051)
I_R						
$I_{R_{NA}} \times I_{2014+}$				1.053***	1.040***	1.009***
				(0.019)	(0.027)	(0.039)
$I_{R_{EU}} \times I_{2014+}$				1.044***	1.114***	1.035***
				(0.019)	(0.029)	(0.044)
$I_{R_{AA}} \times I_{2014+}$				1.274***	1.081***	1.210***
				(0.024)	(0.048)	(0.075)
N	602,599	602,599	430,801	602,599	602,599	430,801

Exponentiated coefficients
Standard errors in parentheses
*p<0.1; **p<0.05; ***p<0.01

Table S3:

Table S3 is build on six regression models like Table S1, except now we considering neighboring topics or disciplines [CIP {1,7} with CIP {8,9}; SA {1,4}with SA {5, 6}] instead of considering all as previous.

The first model predicts cross topic association using the variables Yp, Zp, Kp, Wp, region count and number of discipline.

The second model predicts cross disciplinary association using the same variables except SA and replacing CIP.

The third model predicts the combination of SA and CIP.

The fourth model is similar to model-1 with additional consideration of effects publications year after 2013, and regional impact.

The model five is similar to mode-2 and model-6 is similar to model-3.

The models are created using glm library of R. All values in the table are generated by feeding the models using stargazer library.

	Dependent variable:					
	X_{SA}	X_{CIP}	$X_{SA\&CIP}$	X_{SA}	X_{CIP}	$X_{SA\&CIP}$
	(1)	(2)	(3)	(4)	(5)	(6)
y	1.033***	1.017***	1.043***	1.036***	1.035***	1.071***
	(0.0004)	(0.001)	(0.002)	(0.0005)	(0.002)	(0.003)
z'_j	0.635***	1.210***	0.838***	0.624***	1.127***	0.750***
	(0.006)	(0.019)	(0.030)	(0.006)	(0.019)	(0.031)
ln k	0.867***	1.740***	1.289***	0.879***	1.861***	1.403***
	(0.005)	(0.016)	(0.026)	(0.005)	(0.016)	(0.026)
ln w	2.258***	0.918***	2.584***	2.261***	0.894***	2.496***
	(0.008)	(0.028)	(0.041)	(0.008)	(0.028)	(0.041)
N_R	1.107***	4.594***	5.094***	0.986***	1.652***	1.924***
	(0.008)	(0.015)	(0.023)	(0.017)	(0.034)	(0.053)
N_{CIP}	1.181***			1.169***		
	(0.007)			(0.007)		
N_{SA}		1.183***			1.171***	
		(0.010)			(0.010)	
I_{2014+}				0.871***	0.735***	0.648***
				(0.016)	(0.030)	(0.047)
$I_{R_{NA}}$				0.872***	0.378***	0.450***
				(0.020)	(0.044)	(0.067)
$I_{R_{EU}}$				0.894***	0.123**	0.132
				(0.020)	(0.055)	(0.085)
$I_{R_{AA}}$				0.725***	0.188***	0.130
				(0.021)	(0.057)	(0.096)
I_R						
$I_{R_{NA}} \times I_{2014+}$				1.063***	0.860***	0.842***
				(0.018)	(0.048)	(0.072)
$I_{R_{EU}} \times I_{2014+}$				1.031***	1.228***	1.039***
				(0.019)	(0.069)	(0.113)
$I_{R_{AA}} \times I_{2014+}$				1.118***	0.646***	0.711***
				(0.021)	(0.092)	(0.160)
N	602,599	602,599	396,471	602,599	602,599	396,471

Exponentiated coefficients
Standard errors in parentheses
*p<0.1; **p<0.05; ***p<0.01

Table S4:

Table S4 is based on six regression models. In these models we are measuring effects of different variables on citations like effects of number of co-authors(Kp), Major Mesh Count(Wp,) scholars age(tau) and presence of different year, topic and discipline as constant.

The models are created using plm library of R. All values in the table are generated by feeding the models using stargazer library.

	Dependent variable:					
	Z_p (1)	Z_p (2)	Z_p (3)	Z_p (4)	Z_p (5)	Z_p (6)
ln k	0.413*** (0.002)	0.413*** (0.002)	0.419*** (0.002)	0.434*** (0.003)	0.423*** (0.002)	0.404*** (0.002)
ln w	0.036*** (0.003)	0.041*** (0.003)	0.042*** (0.003)	0.028*** (0.005)	0.056*** (0.004)	0.042*** (0.004)
τ	−0.011*** (0.0001)	−0.011*** (0.0001)	−0.011*** (0.0001)	−0.013*** (0.0002)	−0.010*** (0.0002)	−0.010*** (0.0002)
$I_{X_{SA}}$	0.049*** (0.003)					
$I_{X_{CIP}}$	0.073*** (0.003)					
$I_{X_{Neighboring,SA}}$		0.089*** (0.003)				
$I_{X_{Neighboring,CIP}}$		0.070*** (0.003)				
$I_{X_{Distant,SA}}$			−0.009*** (0.003)			
$I_{X_{Distant,CIP}}$			0.023*** (0.006)			
$I_{X_{SA\&CIP}}$				0.139*** (0.005)		
$I_{X_{Neighboring,SA\&CIP}}$					0.134*** (0.006)	
$I_{X_{Distant,SA\&CIP}}$						0.045*** (0.011)
N	824,306	824,306	824,306	357,859	551,771	526,904
R^2	0.101	0.102	0.100	0.130	0.090	0.092
Adjusted R^2	0.092	0.092	0.091	0.109	0.076	0.077
Standard errors in parentheses:						*p<0.1; **p<0.05; ***p<0.01

Table S5:

Table S5 is based on three regression models. In these models we are measuring effects of different variables on citations like effects of number of co-authors(kp), Major Mesh Count (Wp), scholars age(tau,) publications after 2014 and presence of different year, topic and discipline as constant.

The first model considers the topics and disciplines and interaction with the publications after 2014.

The second model considers co-occurring for neighboring cross topics and disciplines and interaction with the publications after 2014.

The third model considers co-occurring distant cross topics and disciplines and interaction with the publications after 2014.

The models are created using plm library of R. All values in the table are generated by feeding the models using stargazer library.

	Dependent variable:					
	Z_p (1)	Z_p (2)	Z_p (3)	Z_p (4)	Z_p (5)	Z_p (6)
ln k	0.413*** (0.002)	0.413*** (0.002)	0.419*** (0.002)	0.434*** (0.003)	0.423*** (0.002)	0.404*** (0.002)
ln w	0.036*** (0.003)	0.041*** (0.003)	0.042*** (0.003)	0.028*** (0.005)	0.056*** (0.004)	0.042*** (0.004)
τ	−0.011*** (0.0001)	−0.011*** (0.0001)	−0.011*** (0.0001)	−0.013*** (0.0002)	−0.010*** (0.0002)	−0.010*** (0.0002)
$I_{X_{SA}}$	0.049*** (0.003)					
$I_{X_{CIP}}$	0.073*** (0.003)					
$I_{X_{Neighboring,SA}}$		0.089*** (0.003)				
$I_{X_{Neighboring,CIP}}$		0.070*** (0.003)				
$I_{X_{Distant,SA}}$			−0.009*** (0.003)			
$I_{X_{Distant,CIP}}$			0.023*** (0.006)			
$I_{X_{SA\&CIP}}$				0.139*** (0.005)		
$I_{X_{Neighboring,SA\&CIP}}$					0.134*** (0.006)	
$I_{X_{Distant,SA\&CIP}}$						0.045*** (0.011)
N	824,306	824,306	824,306	357,859	551,771	526,904
R^2	0.101	0.102	0.100	0.130	0.090	0.092
Adjusted R^2	0.092	0.092	0.091	0.109	0.076	0.077
Standard errors in parentheses:						*p<0.1; **p<0.05; ***p<0.01

Figure 5:

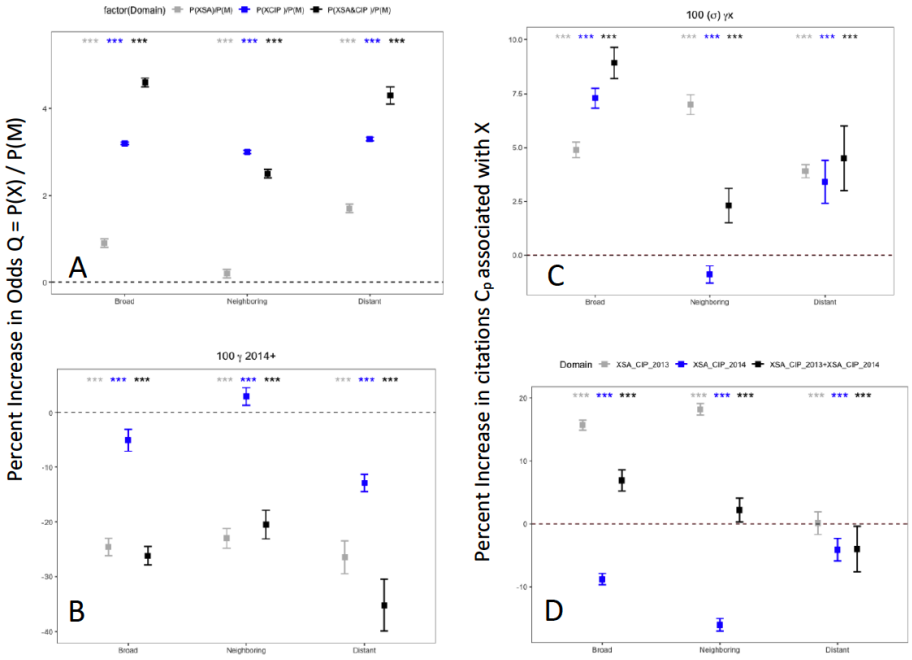

Figure 5(A) represents the increase in odds for topics(SA), disciplines(CIP) or both. It considers the

coefficients of the variable Yp by 100 times and visualize the confidence interval from table S1, S2, S3. Here we can see cross topic has high impact for broad category, but cross discipline has for neighboring categories.

Figure 5(B) represents the increase in odds for topics(SA), disciplines(CIP) or both. It considers the coefficient of the variable Yp after 2014 by100 times and also visualize the confidence interval from table S1, S2, S3.

Figure 5(C) represents the increase in Cp for topics(SA), disciplines(CIP) or both. It considers the co-officient of SA, CIP or both, Neighbouring SA, Neighbouring CIP or both, Distant SA, Distant CIP or both from table S4.

Figure 5(D) represents the increase in Cp for Publication after 2014, topics(SA) & disciplines(CIP) and their interaction. It considers the coefficient of publications after 2014, co-occurrence SA, CIP, their interactions; , co-occurrence neighbouring-SA, neighbouring-CIP and their intercation with publications after 2014 and co-occurrence Distant-SA, Distant-CIP & their interactions with publications after 2014 from table 5.