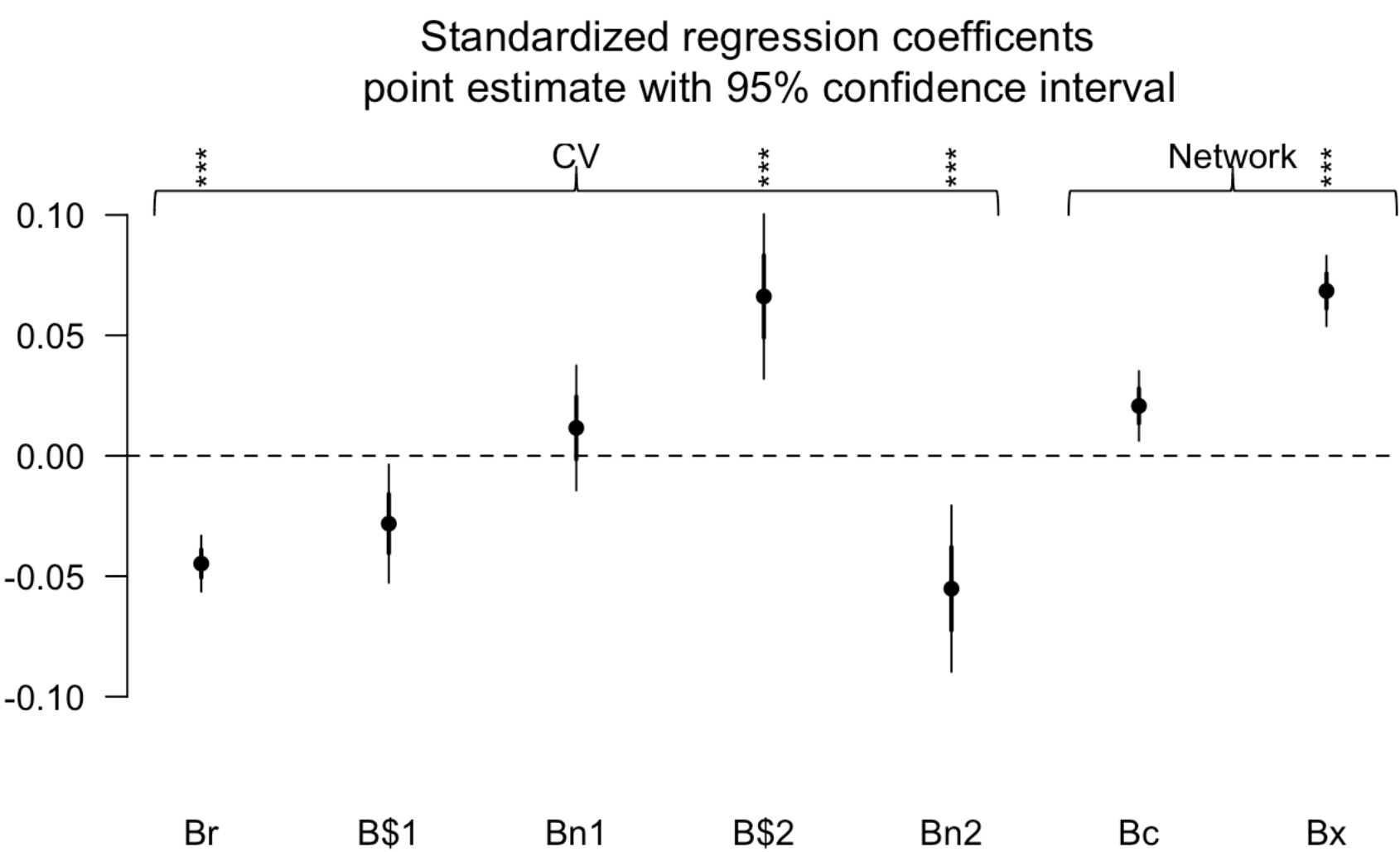# COSC6323: Group Assignment. Part II.

Reproducing figures 'Cross-disciplinary evolution of the genomics revolution'

Vitalii Zhukov, Eljose E Sajan, Jonathan Plata

4/4/2019

Fig 4 Career cross-sectional regression model.

Comments: OLS parameter estimates for the linear regression model in Eq. 1. The coefficients for the relevant covariates split into two categories are shown, depending on whether you might find the information in the researcher's CV or by analyzing her/his collaboration network. To facilitate comparison of the relative strength of the parameter estimates, the standardized beta coefficients are shown, representing the change in the dependent variable ln C_i that corresponds to a 1-SD shift in a given covariate. See table S2 for the complete list of parameter estimates. The levels of statistical significance are as follows: ***$P <= 0.001$.

Conclusions:

Table S2. Career data set: Pooled cross-sectional model.

Predictors	CV			CV + Network			CV + Network [Standardized]		
	Estimates	std. Error	p	Estimates	std. Error	p	Estimates	std. Error	p
Departmental rank ,Br	-0.052	0.007	**<0.001**	-0.047	0.007	**<0.001**	-0.045	0.006	**<0.001**
Productivity (h-index), Bp	1.857	0.021	**<0.001**	1.866	0.025	**<0.001**	0.939	0.009	**<0.001**
Total NSF funding, B$1	-0.005	0.002	**0.020**	-0.005	0.002	**0.027**	-0.028	0.012	**0.022**
# of NSF grants, Bn1	0.024	0.014	0.085	0.013	0.014	0.362	0.012	0.013	0.373
Total NIH funding, B$2	0.016	0.003	**<0.001**	0.014	0.003	**<0.001**	0.066	0.017	**<0.001**
# of NIH grants, Bn2	-0.067	0.018	**<0.001**	-0.061	0.018	**0.001**	-0.055	0.017	**0.001**
PageRank centrality, Bc				0.041	0.015	**0.006**	0.021	0.007	**0.004**
Cross-disciplinarity, Bx				0.571	0.077	**<0.001**	0.068	0.007	**<0.001**
Constant	1.398	0.388	**<0.001**	1.706	0.558	**0.002**	-0.487	0.172	**0.005**
Observations	4190			3900			3900		
R² / adjusted R²	0.884 / 0.883			0.883 / 0.882			0.883 / 0.882		

Comments: The dependent variable is career achievement, measured as the natural logarithm of the Google Scholar citations, LnCi as of 2017. The regression model is specified in Eq. (1) and estimated using standard OLS; there are 4,190 Fi (observations) for the pure CV model and 3,900 observations for the other two models that include network attributes, as in these cases we exclude from consideration disconnected Finodes. Natural logs were used to obtain variables that are approximately normally distributed. Thus, when the independent variable enters in ln, then B corresponds to the % change in Ci following a 1% change in the independent variable; in the case of the cross-disciplinarity fraction, Bx represents the % change in Ci following a 0.01 shift increase in Xi. The first column cluster shows the estimates using only standard CV variables. The combined CV + Network model demonstrates that Fi with larger Xi correlate with higher net citation impact. For the combined model we also report the standardized beta coefficients – useful for comparing the relative strength of covariates within the regression. Standard errors were calculated using the clustered sandwich estimator, clustering on F age-cohort y0 (based on 14 non-i i,5 overlapping 5-year career birth year groups, e.g., 1940-1944, 1945-1950, etc.) to account for within-age- cohort correlation. Also additional fixed effects were included in the regression model: Discipline (O) dummy, 5-year cohort (y0i,5) dummy.

Conclusions:

Table S3.Career data set: Pooled cross-sectional model—robustness check.

Predictors	A. [PR] Estimates	std. Error	p	B. [B] Estimates	std. Error	p	C. [D] Estimates	std. Error	p	D. [-Bn1,-Bn2] Estimates	std. Error	p	E. [-Br] Estimates	std. Error	p
Departmental rank, Br	-0.047	0.007	**<0.001**	-0.042	0.007	**<0.001**	-0.044	0.007	**<0.001**	-0.046	0.007	**<0.001**			
Productivity (h-index), Bp	1.866	0.025	**<0.001**	1.901	0.025	**<0.001**	1.849	0.024	**<0.001**	1.862	0.024	**<0.001**	1.893	0.024	**<0.001**
Total NSF funding, B$1	-0.005	0.002	**0.027**	-0.004	0.002	**0.043**	-0.005	0.002	**0.013**	-0.003	0.001	**0.011**	-0.005	0.002	**0.014**
# of NSF grants, Bn1	0.013	0.014	0.362	0.011	0.014	0.464	0.005	0.014	0.702				0.005	0.014	0.731
Total NIH funding, B$2	0.014	0.003	**<0.001**	0.012	0.003	**<0.001**	0.012	0.003	**<0.001**	0.003	0.001	**0.018**	0.012	0.003	**<0.001**
# of NIH grants, Bn2	-0.061	0.018	**0.001**	-0.069	0.018	**<0.001**	-0.067	0.017	**<0.001**				-0.063	0.017	**<0.001**
PageRank centrality, Bc	0.041	0.015	**0.006**							0.042	0.015	**0.005**	0.057	0.015	**<0.001**
Betweeness centrality, Bb				-0.000	0.006	0.973									
Degree centrality, Bd							0.052	0.010	**<0.001**						
Cross-disciplinarity, Bx	0.571	0.077	**<0.001**	0.557	0.077	**<0.001**	0.524	0.077	**<0.001**	0.579	0.076	**<0.001**	0.549	0.077	**<0.001**
Constant	1.706	0.558	**0.002**	1.238	0.517	**0.017**	1.382	0.535	**0.010**	1.711	0.551	**0.002**	1.655	0.518	**0.001**
Observations	3900			3387			3900			3900			3900		
R² / adjusted R²	0.883 / 0.882			0.874 / 0.873			0.884 / 0.883			0.883 / 0.882			0.881 / 0.881		

Comments: Parameter estimates for variants of the 'CV + Network' pooled cross-sectional models reported in table S2: (a) Model with PageRank centrality. (b) Model with betweenness centrality. (c) Model with degree centrality; (d) Model without the number of grants variables; (e) Model without the departmental rank variable. Results are not significantly different with respect to the primary covariate of interest, that is, cross-disciplinarity. Also additional fixed effects were included in the regression model: Discipline (O) dummy, 5-year cohort (y0i,5) dummy * p <= .05, ** p<= .01, *** p <= .001

Conclusions: