

for a given unit/time period. For example, if a researcher is interested in whether district-level rainfall (in the prior month) predicts conflict, (s)he may decide to study monthly rainfall and conflict levels for 100 countries from 1950-2000. Here, the researcher might choose to include fixed effects for each country (that is, including 99 control variables, one for each country except a reference country), to control for the ‘average’ rate of conflict in each of the 100 countries studied between 1950-2000. The regression results will identify the impact of recent rainfall on local conflict, controlling for average levels of conflict in each country. So the regression will tell us: how much variation in local conflict *within each country (that is, deviation from the country-specific baseline level of conflict)* is explained by (district-level recent) rainfall levels?

1 f) **(5 Points)** Do you think multicollinearity would be a relevant problem for the main regression model, as implemented in Column 2? How would you know? What would you do to resolve this if it is indeed a problem?

1 g) **(15 Points)** Based on the results in Table 1, do you think that Chinese tariffs are likely to have affected the 2020 US Presidential elections? Explain why (or why not). Finally, write out a regression model that you would use to assess the effectiveness of Chinese tariffs in affecting the electoral outcome: explain what measure you would use for your dependent variable, your main independent variable, and any other factors you consider important to include as controls along with an explanation of why.

Table 1. Counties targeted by tariffs

	Dependent Variable: Workers Employed in Targeted Industries (%)							
	(1) April	(2)	(3) June	(4)	(5) August	(6)	(7) September	(8)
GOP VOTE SHARE (%)	0.018*** (0.004)	0.012*** (0.004)	0.034*** (0.005)	0.028*** (0.005)	0.019*** (0.003)	0.019*** (0.003)	0.136*** (0.008)	0.134*** (0.009)
SWING DISTRICT	-0.035 (0.257)	-2.520*** (0.845)	-0.041 (0.291)	-2.539*** (0.904)	-0.238** (0.105)	0.033 (0.312)	-0.351 (0.427)	-1.051 (1.261)
GOP VOTE SHARE (%) * SWING DISTRICT		0.055*** (0.017)		0.055*** (0.019)		-0.006 (0.006)		0.015 (0.026)
Observations	3,106	3,106	3,106	3,106	3,106	3,106	3,106	3,106

Notes: Dependent variables in models 1 to 2 are the percentage of workers employed in industries targeted by the first round of Chinese retaliation in April. The next models focus on the second, third, and last round of retaliation in June, August and September, respectively. All models include state fixed effects. All regressions are weighted by counties' total voting age population in 2016. Robust standard errors in parentheses. * $p < .10$; ** $p < .05$; *** $p < .01$.

Problem 2 (70 Points):

To answer this question, first download the data set “final_IR602_Spring2021_data.dta” from the Final Exam folder on our course’s Blackboard site.

To what extent does one’s status as the main income earner (the “breadwinner”) limit political ambition for American women? To answer this question, take a look at the data set provided, which studies whether or not women decide to run for office out of a large number who completed a training program to help them compete in and win local electoral races in the US.

2.a) **(5 Points)** Calculate summary statistics on how much the decision to run for office (“ran”, where those who decided to run are coded as 1) varies by breadwinner status (“breadwinner”, where those contributing over 50% of the household income are coded as “breadwinners”, e.g. 1). To do this, utilize the tabstat command, where you can run the command: `tabstat y, by(x)` to see how an outcome variable (y) varies by level of an independent variable (x). Remember, if you want to look at specific statistics, such as mean, standard deviation, and number of observations, you could specify: `tabstat y, by (x) stats(mean sd n)`.