

2.b) **(5 Points)** Write a bivariate regression model you would use to study the relationship between breadwinner status and the decision to run for office, using the basic format: $y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$.

2.c) **(5 Points)** Run the regression in Stata. Interpret the main coefficient of interest.

2.d) **(15 Points)** The authors consider a more nuanced measure of women's economic status in the household based on a finer-grained coding of their economic contribution as either 0-25% of household income, 25-50%, 50-75%, or 75-100% of the household income ("hhexpenses"). Do you think this is likely a better or worse measure of the relationship between women's economic status within the household and political ambition? Why? Answer, then run this bivariate regression and interpret the coefficient on the result. Is this what you expected? Why/not?

2.e) **(15 Points)** What is a likely source of omitted variable bias? Why? What would be the direction of bias you would predict? Take a look at the variables within the dataset and figure out the closest version of this variable available. Write out the equation for the multivariate regression that may solve this form of omitted variable bias. Run the regression. What do you find? Does it present evidence in favor of or contrary to your hypothesis?

2.f) **(15 Points)** Consider the two key measures of "goodness of fit" for the regression models in parts 2.c), 2.d), and 2.e) (that is, the R-squared and Root MSE). What do these look like for each specification? How do the three specifications compare, specifically which model provides the "best" fit?

2.g) **(10 Points)** Considering the regression models you ran in parts 2.c), 2.d), and 2.e), what would you conclude about the relationship between women's economic status within the household and their political ambition? Why?