**Problem 1 [40 points]** Belenzon, Chatterji and Daley (2017) find that eponymy—firms being named after their owners—is linked to firm performance. The baseline specification in their paper takes the form

where the indices and denote firms and years, is return on assets, is a binary indicator variable that takes the value of one if firm ’s name in year includes the name of owner and zero otherwise, and explanatory variables include time-varying controls for log of assets, log number of shareholders, equity dispersion, and is a regression error term. The variables are complete sets of three-digit standard industrial classification (SIC) codes, country, and year dummies. The total number of observations is more than 6 million and there are more than 1 million firms in the data for the period 2002-2012.

(a) [5 points] Belenzon, Chatterji and Daley (2017) propose a signaling explanation for the eponymy-performance relationship. They argue that eponymy creates a stronger association between the entrepreneur and her firm that increases the reputational benefits of having the market hold a favorable impression of the quality of her firm. Consequently, they infer that high-ability entrepreneurs are more drawn to eponymy than are low-ability ones. Based on this explanation, what is your prediction on ? Explain briefly.

(b) [5 points] Given the information given above, do the authors of the paper have balanced panel data? Explain briefly.

(c) [5 points] In the basic specification, there is no firm fixed effect. In the data, only 0.3 percent of firms change their eponymy status during the sample period. Can you conjecture why the authors did not include the firm fixed effect in their basic specification? Justify your answer.

(d) [5 points] Suppose that you would like to control for the owner fixed effect, in addition to all the controls in the basic specification. In this case, what type of variation in do you exploit? In the data, the same person may own multiple firms. [2 points]

(e) [5 points] Using the dataset **EPO\_AER\_subsample.dta**, which is a five percent random sample of the dataset used in Belenzon, Chatterji and Daley (2017), estimate the model given in part (a). Use the areg command in STATA since it is necessary to deal with many indicator variables for the SIC codes. Use the standard error clustered by firm. Discuss your estimation results.

(f) [5 points] Using only observations for year 2007, re-estimate the same model above. Compare your estimation results and provide brief remarks.

(g) [5 points] Now include the firm fixed effect and estimate the regression model. Use the xtset, xtreg commands in STATA to run panel data regression. Use the standard error clustered by firm. Discuss your findings.

(h) [5 points] Restrict the sample such that owners have at least two different businesses. Using this subsample and include the owner-fixed effect to estimate the model. Use the standard error clustered by firm. Discuss your findings.