Complete the tasks below:

1. Using appropriate data sources (e.g. Compustat, CRSP, and BoardEx)1, download data (see below) for at least 50 US firms for the period 2010-20192,3 . Present a table of variable definitions. Label this table: Table1: Variable Definitions. Make sure to include all the variables used in the project4 . (1000 words max) [10 marks]

2. Present summary statistics for all the variables used in the project. Make sure to include: mean, standard deviation, min, max, 25% percentile, 50% percentile, 75% percentile, number of firms, number of firm-year observations. Label this table: Table2: Summary Statistics5. Make sure to winsorise variables when appropriate and in these cases include both, the summary statistics for the raw variable and for the winsorised one. This table does not count towards the word count. [5 marks]

3. Collapse the data into a cross-sectional sample; i.e., compute time averages for every variable for each firm. You want to assess whether firms with stronger corporate governance take less risk than firms with weaker corporate governance. Based on a strong literature review and data availability, choose one proxy for firm risk and at least 2 variables relevant to the governance of firms.6 One of these proxies should be used as a dummy variable, so make a sensible choice. Control for firm size. Using all these variables, run a cross-sectional regression and report the most appropriate standard errors. Label this table: Table 3: Cross-sectional Regression. This table does not count towards the word count. Clearly explain your . rationale for choosing these standard errors (feel free to include additional tables if required) and write down the STATA command used to compute the selected standard errors (as a footnote for this Table). (50 words max) [10 marks]

1 A video to help you merge databases has been uploaded in the course’s Learn webpage. 2 The choice of companies can demonstrate one’s approach and curiosity in answering interesting research questions in finance. For example, you could focus on a particular industry: oil&gas, renewables, IT, media content creation, fintech, electric cars, etc.; or you could choose other interesting criteria, for example the firms with the largest short interest positions, or with the highest paid CEOs, or with the most women on their boards, or the highest ranked by their employees, etc. 3 Most likely you will have an unbalanced sample. This is okay! Just make sure you have at least 5 years of data available for each firm. 4 An example of a variable definitions table is included at the end of the document. 5 Note that sometimes zeroes denote missing values. Make sure your summary statistics look sensible. Marks will be deducted for wrong construction of variables or careless calculations. 6 You should think about these variables at the same time as when you think about your sample. One such consideration is to make sure that there is enough variation in all your variables. For instance, if your sample is the firms that have the highest proportion of female directors on their board, then the proportion of female directors in the boardrooms will not have enough variation to explore the relationship of interest. Therefore, this will not be a good explanatory variable

4. Rather than using a cross-sectional sample, you will now be using panel data. Add time dummies to the specification in Table 3. Run a firm-fixed effects regression and include this table in your project. Label this table: Table 4: Fixed Effects. Make sure you use cluster-robust standard errors (Stata option -vce(cluster varname)-). Include your Stata regression command (as a footnote for this Table). This table does not count towards the word count. [10 marks]

5. Add an interaction term between the 2 corporate governance variables to the model reported in Table 4. Label this table: Table 5: Interaction Effects. This table does not count towards the word count. Carry out an analysis using the margins command. Explain the effect of one of your governance variables on firm risk. Are these effects statistically significant? Make sure to include relevant tables and plots that may add value to your answer. (100 words max) [10 marks]

6. Provide an economic interpretation of the interaction term. (50 words max) [10 marks]

7. Is the model presented in Table 5 misspecified? Answer Yes or No. Show tables and results from any statistical test that you are performing to support your answer. (1 word max) [10 marks]

8. Compare the results of the models in Table 4 and Table 5: the magnitudes and significance of the coefficients. Come up with at least two plausible explanations for some of the differences across models. (100 words max) [10 marks]

9. Give an example of an unobserved omitted variable that could lead to one of your corporate governance estimated coefficients being biased. Justify your choice. (100 words max) [10 marks]

10. Your governance variables are endogenous. The election of Donald Trump was unexpected. Explain whether using this shock would allow you to carry out a Diff-indiff analysis that satisfies the required assumptions and has cleaner causal interpretation than OLS (answer Yes or No) and why. (100 words max) [10 marks]

11. Choose the largest firm in your sample. Download its daily prices and assess whether your price series is stationary and/or weakly dependent. Make sure to answer Yes or No and briefly support your answer (e.g. you could use a graph or a formula). (50 words max)