

FIN9013 Assignment 1

Ang Zhang

February 4, 2025

Linear regression, interactions, split-samples and non-linearities

In the baseline model, the slope term of *Size* factor is 0.0237, which means that for a one unit increase in the logarithm of annual sales turnover, measured in millions, the predicted value of the debt ratio, measured by the total debt over total book asset, increase by 0.0237, holding all other variables constant. The slope term of *MB* factor is 0.0085, which means that for a one unit increase in the market to book ratio of assets, the predicted value of the debt ratio, measured by the total debt over total book asset, increase by 0.0085, holding all other variables constant. The slope term of $D(RD)$ factor is -0.0902, which means that for a firm with positive R&D spending, the predicted value of the debt ratio, measured by the total debt over total book asset, is 0.0902 lower than a comparable firm with no R&D spending, holding all other variables constant.

However, we can hardly say that this model is a good fit, as the R^2 value is only 0.0813, which means that only 8.13% of the variation in the dependent variable, *Debt*, can be explained by the independent variables, *Size*, *MB* and $D(RD)$, while the rest 91.87% of the variation is due to other factors not included in the model.

When all variables are held at their mean, the predicted value of *Debt* is 0.2029, while for all variables at their median the predicted *Debt* value is 0.1655.

The effect of $D(RD)$ on *Debt* is constant and always at β_{rd} , no matter where MB stands at. That's because in eq.(1), there's no interaction term so the coefficient of $D(RD)$ which is β_{rd} solely determines the effect that $D(RD)$ has on *Debt*.

Comparing the coefficient of *MB* in the RD Model and the No RD Model, we can see that the coefficient of *MB* in the RD Model is 0.0015, while in the No RD Model it is -0.0173. This suggests that the effect of *MB* have different effects on the debt ratio for firms with

and without R&D spending.

Comparing the two models where RD *rank* is included as rank in one of them (RD Rank Model) and RD *rank* is included as class in the other one (RD Class Model), we can see that the parameter estimates for the four categorical variables are quite different. Also, the RD Rank Model has a higher R^2 value, which means that it explains more of the variation in the dependent variable. This suggests that the effect of R&D spending might be likely non-linear, and that the RD Rank Model is a better fit for the data.

Discrete choice variables and censoring

Instrumental variables

Appendix

Table 1: Summary Statistics Table

Panel A: All Firms								
Variable	N	Mean	StdDev	Min	P25	Median	P75	Max
Debt	16257	0.197	0.225	0	0.000	0.129	0.317	1.017
Size	16257	5.786	2.060	-0.997	4.576	5.991	7.194	10.068
MB	16257	1.912	1.549	0.334	0.923	1.397	2.344	8.834
D_RD	16257	0.582	0.493	0	0	1.000	1.000	1.000
RD_rank	16257	1.908	1.120	1.000	1.000	2.000	2.000	5.000
D_Debt	16257	0.336	0.472	0	0	0	1.000	1.000
D_DecIPO	16257	0.089	0.285	0	0	0	0	1.000
Panel B: Firms with R&D								
Variable	N	Mean	StdDev	Min	P25	Median	P75	Max
Debt	9468	0.150	0.201	0	0	0.058	0.246	1.017
Size	9468	5.183	2.168	-0.997	3.908	5.299	6.658	10.068
MB	9468	2.202	1.699	0.334	1.064	1.644	2.760	8.834
D_RD	9468	1.000	0	1.000	1.000	1.000	1.000	1.000
RD_rank	9468	2.559	1.068	2.000	2.000	2.000	3.000	5.000
D_Debt	9468	0.245	0.430	0	0	0	0	1.000
D_DecIPO	9468	0.079	0.270	0	0	0	0	1.000
Panel C: Firms without R&D								
Variable	N	Mean	StdDev	Min	P25	Median	P75	Max
Debt	6789	0.264	0.239	0	0.049	0.226	0.401	1.017
Size	6789	6.626	1.548	-0.997	5.686	6.647	7.689	10.068
MB	6789	1.508	1.199	0.334	0.807	1.140	1.748	8.834
D_RD	6789	0	0	0	0	0	0	0
RD_rank	6789	1.000	0	1.000	1.000	1.000	1.000	1.000
D_Debt	6789	0.463	0.499	0	0	0	1.000	1.000
D_DecIPO	6789	0.103	0.304	0	0	0	0	1.000

Note: This table presents the summary statistics for the variables used in the analysis. Panel A includes all firms, Panel B includes firms with R&D, and Panel C includes firms without R&D.

Table 2: Parameter Estimates of Each Model

Parameter	Model Baseline	Model Interaction	Model Augmented	Model RD	Model No RD	Model RD Rank	Model RD Class
Intercept	0.1114*** (0.0068)	0.1298*** (0.0072)	0.1029*** (0.0119)	0.0292*** (0.0063)	0.1029*** (0.0131)	0.0376*** (0.0090)	0.1070*** (0.0071)
Size	0.0237*** (0.0009)	0.0242*** (0.0009)	0.0282*** (0.0017)	0.0227*** (0.0009)	0.0282*** (0.0018)	0.0297*** (0.0010)	0.0334*** (0.0010)
MB	-0.0033*** (0.0011)	-0.0175*** (0.0022)	-0.0173*** (0.0022)	0.0015 (0.0012)	-0.0173*** (0.0024)	-0.0072*** (0.0012)	-0.0067*** (0.0011)
D_RD	-0.0771*** (0.0037)	-0.1092*** (0.0055)	-0.0737*** (0.0137)				
MB*D_RD		0.0194*** (0.0025)	0.0188*** (0.0025)				
Size*D_RD			-0.0055*** (0.0020)				
RD_rank						0.0008 (0.0020)	
RD_rank 1							-0.0547*** (0.0081)
RD_rank 2							-0.1358*** (0.0076)
RD_rank 3							-0.1473*** (0.0097)
RD_rank 4							-0.0337** (0.0158)
RSquare	0.1054	0.1086	0.1091	0.0586	0.0419	0.0809	0.1253
#Obs	16257	16257	16257	9468	6789	16257	16257

*Note: This table presents the parameter estimates for various models. The baseline model includes the intercept, Size, and MB. The interaction model adds an interaction term between MB and D_RD. The augmented model includes additional interaction terms. The RD model includes only firms with R&D, while the No RD model includes firms without R&D. The RD Rank model includes the RD_rank variable, and the RD Class model includes categorical variables for RD_rank. Standard errors are in parentheses. Significance levels are indicated by * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*