Problem Set: Matching

Empirical Methods in Corporate Finance

Michael R. Roberts

August 17, 2010

Using data from fiscal year-end 1995, consider estimating the effect of an investment grade credit rating on corporate investment. Define the treatment group as those firms with an investment grade credit rating, and control as those firms with a speculative grade credit rating. Exclude all financial firms and utilities, as well as foreign governments, international affairs non-operating establishments and companies not headquartered in the United States (see attachment).

1. What is the endogeneity concern?

2. Compute sample averages for the treatment and control groups for the following variables: investment, cash flow, market-to-book, log(assets), tangibility, and book leverage. Perform paired sample t-statistics, allowing for different variances across the samples. Discuss the results.

3. Estimate the treatment effect on investment using a multivariate linear regression that controls for the firm characteristics from #1. Try two specifications: (1) only linear controls, and (2) 5th order polynomial for each control. How do your treatment effect estimates compare to each other and to that found in #1.

4. Estimate the average treatment effect (ATE), average treatment effect of the treated (ATT), and average treatment effect of the untreated (ATU) using a propensity score matching estimator. For the matching variables, use the same controls as used in #2. Report the results of the estimated propensity score model, and the frequency counts for the treatment and control groups on and off the common support of propensity scores. Perform a balancing test to determined to what extent the matching eliminated observable differences.

5. Does matching solve the endogeneity concern? Why or why not? What alternative approaches can one take to determine the impact of an investment grade credit rating on corporate investment?

I. Data

The data is from the annual Compustat database, FUNDA, and is located on WRDS at "/wrds/comp/sasdata/na". In the investment and capital structure literatures, a variety of screens are frequently used to eliminate certain observations from the sample. A few of these screens include the following.

1. (indfmt == "INDL" & datafmt == "STD" & popsrc == "D" & consol == "C"): These conditions ensure that gvkey-datadate uniquely identify each observation. (gvkey-fyear is "almost" the unique identifier, but for 48 obs.)
2. (year $\geq$ 1965): Observations with year-ends greater than or equal to 1965. Prior to this year, selection issues become particularly severe in Compustat.
3. (sic $\geq$ 0000 & sic $\leq$ 999): Agriculture, Fishing & Hunting.
4. (sic $\geq$ 4900 & sic $\leq$ 4999): Utilities.
5. (sic $\geq$ 6000 & sic $\leq$ 6999): Financial Firms.
6. (sic = 8888): Foreign Governments.
7. (sic $\geq$ 9000 & sic $\leq$ 9999): International affairs & non-operating establishments.
8. (gvkey != ""): No Missing company indicators.
9. (fyear != .): No Missing fiscal year indicators.
10. (fic == "USA"): Only firms headquartered in the United States.
11. (prcc_f = . & csho = .): Nonmissing stock market data (price & shares outstanding).

The investment variable definitions come from Hennessy, Levy, & Whited 2007 (JFE). The capital structure variable definitions come from Lemmon, Roberts, and Zender (2008) (JF). (In the construction of the market-to-book ratio here, you must first set all missing observations for pstkl and txditc to zero.)

			Investment Variables	
g	mb	=	(at + (prcc_f * csho) - ceq - txdb) / at;	(Market-to-Book)
g	cf	=	(ib + dp);	(Cash flow)
g	cf_k	=	cf / ppent[_n-1];	(Cash flow / capital(t-1))
g	inv	=	(capxv - sppe);	(Net Investment)
g	inv_k	=	inv / ppent[_n-1];	(Investment / capital(t-1))
			Capital Structure Variables	
g	td	=	dlc + dltt;	(Total Debt)
g	bl	=	td / at;	(Book leverage)
g	ml	=	td / (td + (prcc_f * csho));	(Market leverage)
g	prof_a	=	oibdp / at;	(Profitability)
g	tang_a	=	ppent / at;	(Tangibility)
g	me	=	prcc_f * csho;	(Market equity)
g	mk2bk	=	(me + dlc + dltt + pstkl + txditc) / at;	(Market-to-book)
g	loga	=	log(at);	(Log(assets))
g	logsale	=	log(sale);	(Log(sales))
g	zscore	=	(3.3 * pi + sale + 1.4 * re	
		+	1.2 * (act - lct)) / at;	(Altman's unlevered Z-score)