SESSION 5B: MORE ON REGRESSION APPLICATIONS

Session 5
Looking for links!

1. Building a Regression

- While the mechanics of a simple or multiple regression always involve finding independent variables to explain/predict a dependent variable, the process by which you find those variables is important.

- In general, you can find the independent variables statistically by scanning the data and finding variables purely based upon correlations with the dependent variable.
  - While this approach works statistically, it has key problems. If you have enough data and variables to work with, data mining can lead you to variables that provide high R-squared and statistical significance, but there is little you are learning about the past or gaining in terms of predictive value.

- A better approach is to start with intuition and/or a model that links the independent variables to the dependent variable, and then running regressions to see if the model holds. The plus is that if you find statistical significance, you have a much better basis for analysis and prediction. The minus is that the data might not back up your model, either because the model does not hold or because the independent variables that you picked at not good proxies.

An Example: PE Ratio Determinants

- To understand the fundamentals, start with a basic equity discounted cash flow model.
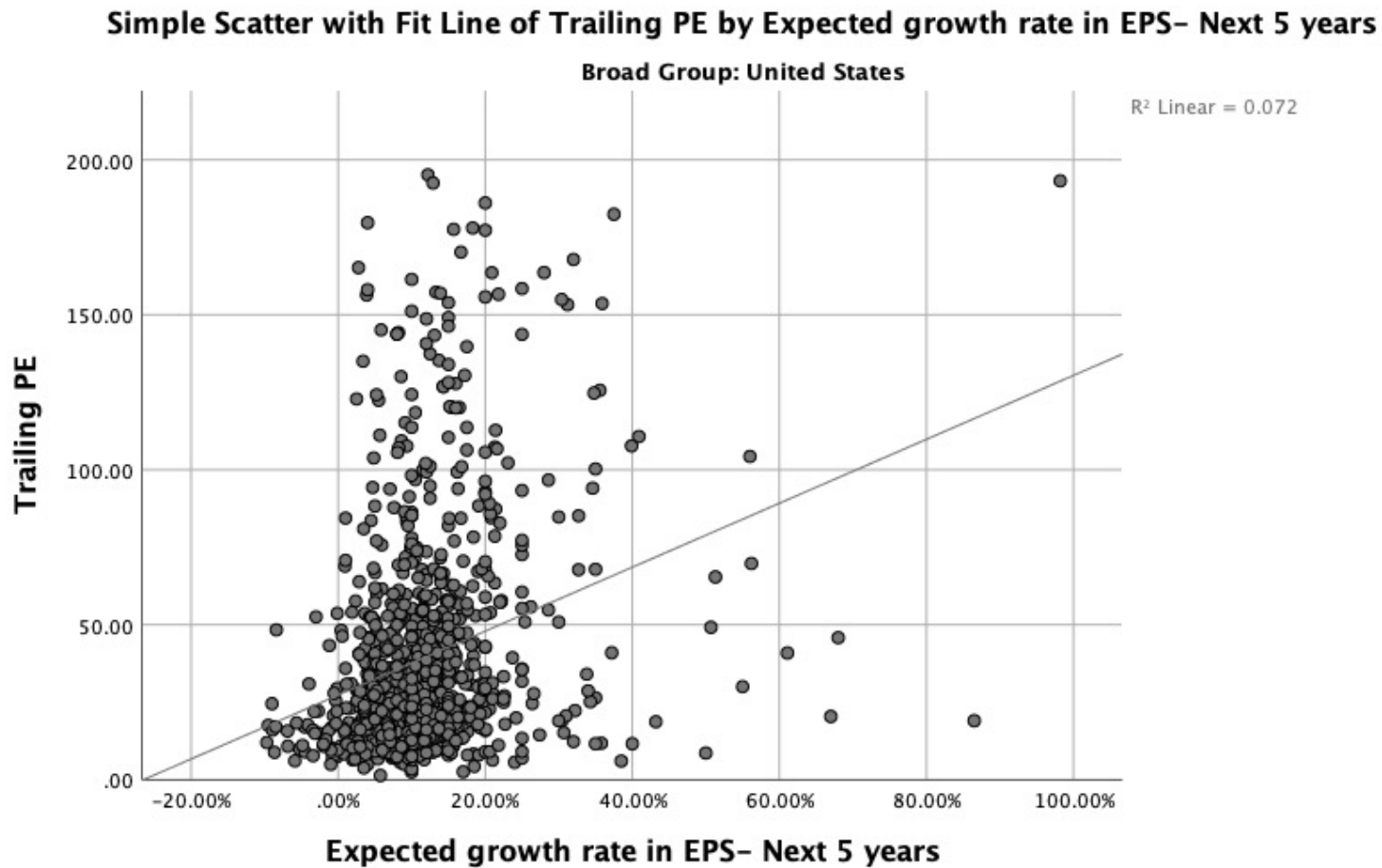 - With the dividend discount model,

  $$P_0 = \frac{DPS_1}{r - g_n}$$

  - Dividing both sides by the current earnings per share,

  $$\frac{P_0}{EPS_0} = PE = \frac{\text{Payout Ratio}*(1 + g_n)}{r - g_n}$$

- If this model is right, the PE ratio for a company should be determined by three variables:
  - Payout ratio, with higher payout ratios leading to higher PE
  - Growth rate, with higher growth rates leading to higher PE
  - Risk, with higher discount rates (r) leading to lower PE

Aswath Damodaran

Checking Linearity: PE versus Expected EPS Growth in January 2021

Simple Scatter with Fit Line of Trailing PE by Expected growth rate in EPS– Next 5 years

Broad Group: United States

R^2 Linear = 0.072

Aswath Damodaran

PE Ratio: Standard Regression for US stocks - January 2021

Model Summary[a]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.629[b]	.396	.394	4035.87822

a. Broad Group = United States

b. Predictors: (Constant), Expected growth rate in EPS–Next 5 years, Beta, Payout ratio

The regression is run with growth and payout entered as absolute, i.e., 25% is entered as 25)

Coefficients[a,b,c]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.104	2.828		1.451	.147
	Payout ratio	.174	.017	.259	10.087	.000
	Beta	1.714	2.709	.015	.633	.527
	Expected growth rate in EPS– Next 5 years	2.304	.087	.681	26.512	.000

a. Broad Group = United States

b. Dependent Variable: Trailing PE

c. Weighted Least Squares Regression – Weighted by Market Cap (in US $)

Aswath Damodaran

Don't fight the data: If a coefficient is not significant, take it out…

Model Summary[a]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.623[b]	.389	.388	4049.88731

a. Broad Group = United States

b. Predictors: (Constant), Expected growth rate in EPS–
Next 5 years, Payout ratio

Coefficients[a,b,c]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.913	1.650		3.584	.000
	Payout ratio	.171	.017	.254	9.921	.000
	Expected growth rate in EPS– Next 5 years	2.284	.087	.674	26.336	.000

a. Broad Group = United States

b. Dependent Variable: Trailing PE

c. Weighted Least Squares Regression – Weighted by Market Cap (in US $)

Aswath Damodaran

If a coefficient has the wrong sign: The Multicollinearity Problem

Correlations[a]

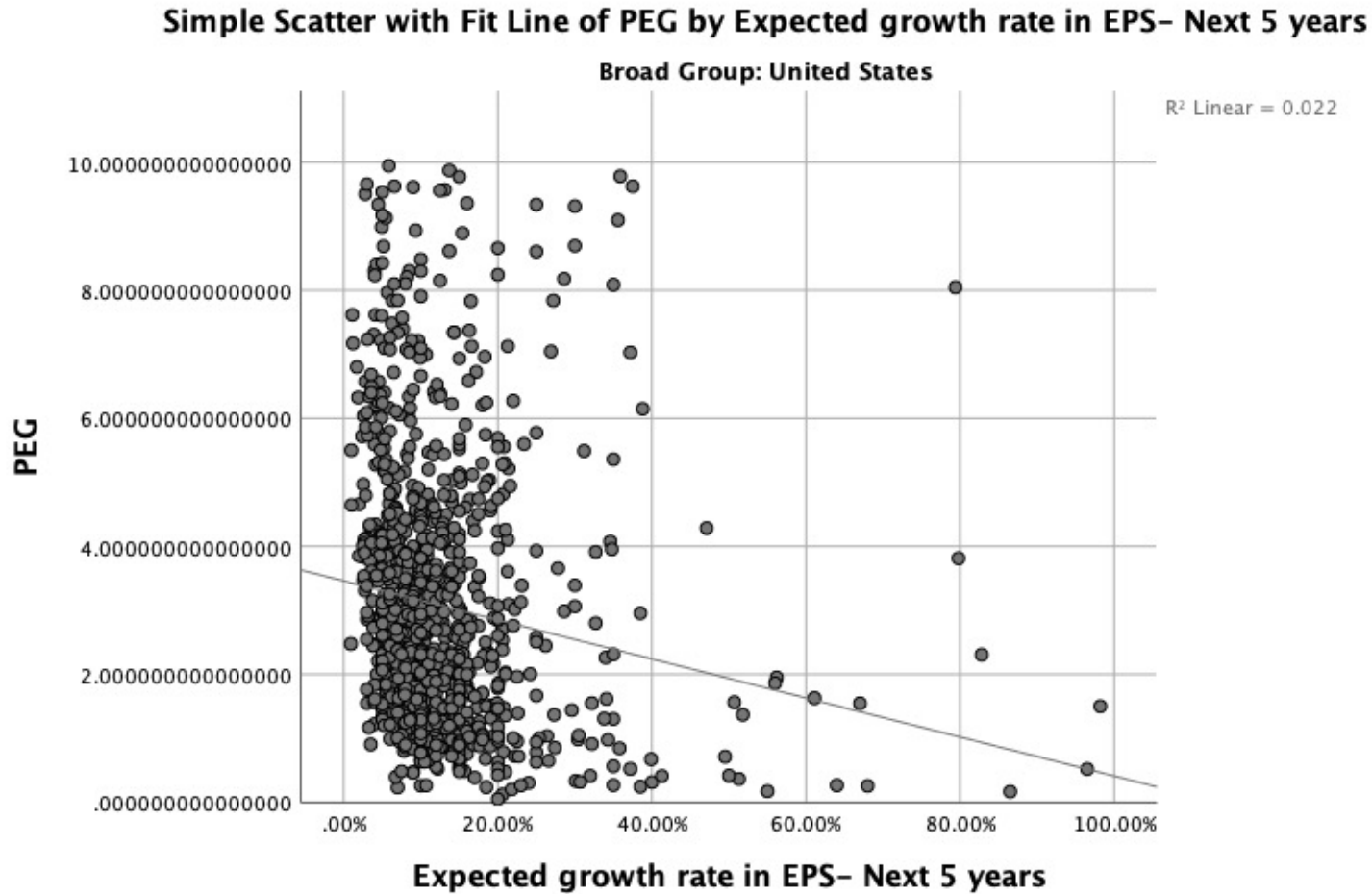|		Trailing PE	Payout ratio	Expected growth rate in EPS– Next 5 years	Beta
Trailing PE	Pearson Correlation	1	.144[**]	.270[**]	.071[**]
	Sig. (2–tailed)		.000	.000	.001
	N	2348	2320	1109	2293
Payout ratio	Pearson Correlation	.144[**]	1	−.220[**]	.080[**]
	Sig. (2–tailed)	.000		.000	.000
	N	2320	2434	1138	2364
Expected growth rate in EPS– Next 5 years	Pearson Correlation	.270[**]	−.220[**]	1	−.093[**]
	Sig. (2–tailed)	.000	.000		.000
	N	1109	1138	1649	1591
Beta	Pearson Correlation	.071[**]	.080[**]	−.093[**]	1
	Sig. (2–tailed)	.001	.000	.000	
	N	2293	2364	1591	6338

Aswath Damodaran

7

Improving R-Squared?

- <u>Look for better proxies</u>: In the regression, the variables used for growth, risk and payout (the expected growth rate in earnings per share, 2-year regression beta and a payout ratio from the most recent year) may not be the best proxies for future values of each.

- <u>Add more independent variables</u>: You can add more independent variables, but in doing so, you should start with a common sense model of why. Otherwise, you run the risk of kitchen sink regressions.

- <u>Review outliers</u>: In some cases, a few extreme outliers can alter your R-squared, even with large samples. Capping the values or removing the observations can help, with the caveat that you are "messing" with the data.

- <u>Change to a weighted least square regression</u>: If there are some observations that you believe should count for more than others in your regression, you can try a weighted least square regression.
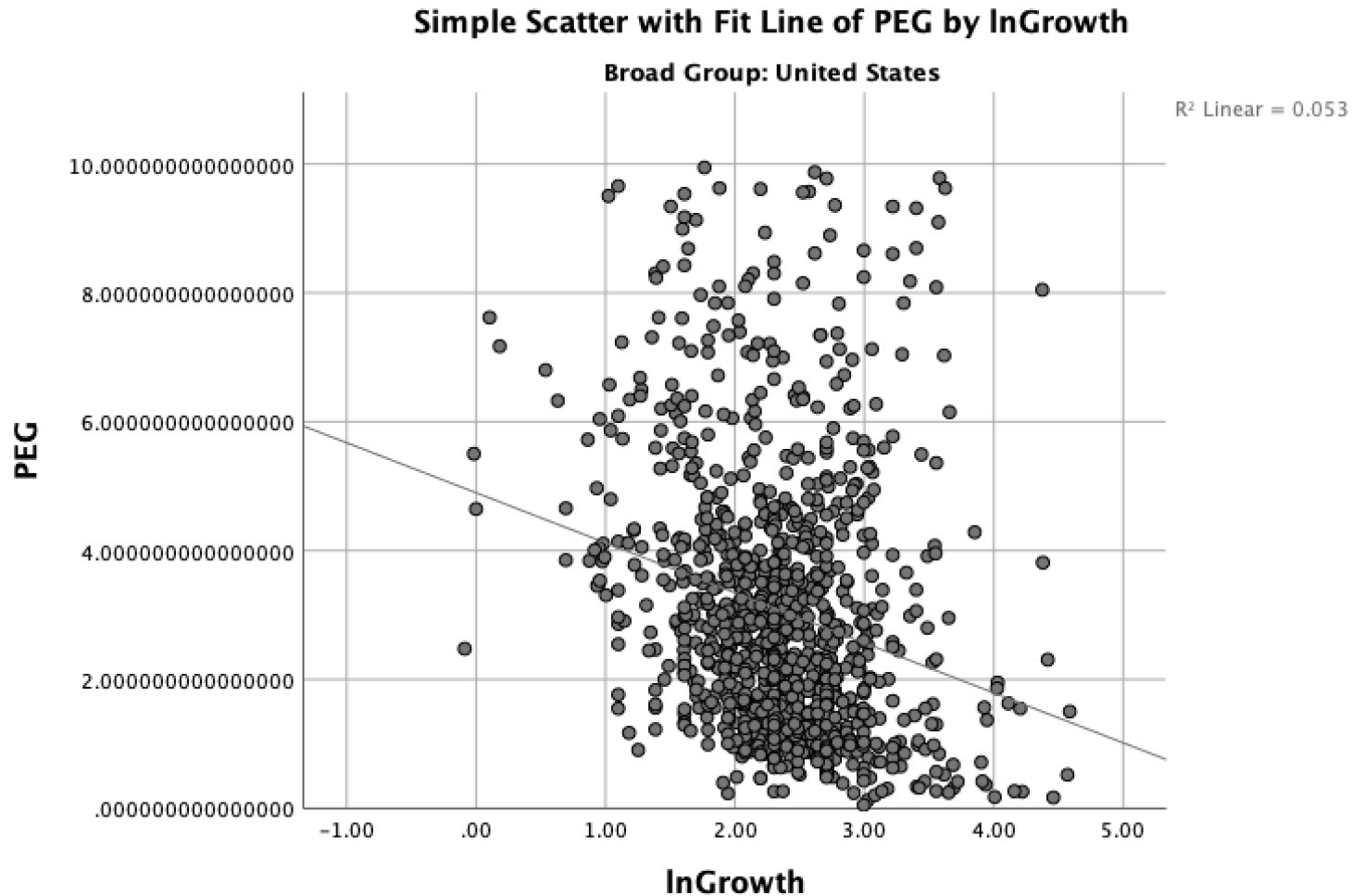
Data Mining, P-hacking and Other Practices…

- **Data mining** refers to the practice of going through data looking for variables that have high explanatory power and presenting this as evidence of co-movement. While there is nothing statistically wrong with this practice, researchers should not attach economic or investing significance to nonsense variables.

- **P-hacking** refers to game playing with the data, changing time periods or variable measurement modes, with the intent of delivering p values and t statistics that are significant.

- Finally, with large samples, **statistical significance does not always translate into economic significance**. There are a lot more ways of making money on paper (based on research and statistical studies) than there are in practice.

2. A Non-linear Relationship: PEG Ratios versus Growth

Simple Scatter with Fit Line of PEG by Expected growth rate in EPS– Next 5 years

Broad Group: United States

R^2 Linear = 0.022

PEG versus ln(Expected Growth)

Simple Scatter with Fit Line of PEG by lnGrowth
Broad Group: United States

Aswath Damodaran

PEG Ratio Regression - US stocks January 2020

Model Summary[a]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.341[b]	.116	.113	1.91045878

a. Broad Group = United States

b. Predictors: (Constant), Beta, Payout ratio, lnGrowth

Coefficients[a,b]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	5.626	.321		17.521	.000
	Payout ratio	.004	.001	.107	3.294	.001
	lnGrowth	−.660	.114	−.190	−5.799	.000
	Beta	−1.138	.170	−.210	−6.696	.000

a. Broad Group = United States

b. Dependent Variable: PEG

Aswath Damodaran

3. Sample sizes and Regressions

- In general, larger sample sizes are better than smaller ones, if you want statistical significance and explanatory power.

- In practice, there are times when you will have only small samples, no matter how hard you try. If you try to run a multiple regression, and use "too many" independent variables, you may get a regression that looks good (on R-squared) but are useless in either explaining or predicting the dependent variable.

- As a rule of thumb,
  - If your sample size is 10-15 observations, the best you can do is run a simple regression
  - For every additional 10-15 observations, you can add one more independent variable.

An Example

Regression Statistics	
Multiple R	0.867415154
R Square	0.752409049
Adjusted R Square	0.504818099
Standard Error	27.78683087
Observations	9

ANOVA

	df	SS	MS	F	Significance F
Regression	4	9385.496876	2346.374219	3.038919828	0.153548553
Residual	4	3088.431879	772.1079698		
Total	8	12473.92876			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-13.47974744	22.33172411	-0.603614274	0.578658552	-75.48255353	48.5230587	-75.482554	48.5230587
EV/Revenues	20.68638521	40.16380231	0.515050469	0.633672769	-90.82620713	132.198978	-90.826207	132.198978
EV/EBITDA	-0.181441851	1.690097701	-0.107355836	0.919675869	-4.873905339	4.51102164	-4.8739053	4.51102164
Price/sales	-11.36003758	31.9251504	-0.355833487	0.739938381	-99.99846515	77.27839	-99.998465	77.27839
Operating Margin	0.55671716	0.734609703	0.757840739	0.490740927	-1.482886354	2.59632067	-1.4828864	2.59632067

4. Dummy Variable Regressions: Telecom ADRs in 1997

Company Name	PE	Growth
PT Indosat ADR	7.8	0.06
Telebras ADR	8.9	0.075
Telecom Corporation of New Zealand ADR	11.2	0.11
Telecom Argentina Stet - France Telecom SA ADR B	12.5	0.08
Hellenic Telecommunication Organization SA ADR	12.8	0.12
Telecomunicaciones de Chile ADR	16.6	0.08
Swisscom AG ADR	18.3	0.11
Asia Satellite Telecom Holdings ADR	19.6	0.16
Portugal Telecom SA ADR	20.8	0.13
Telefonos de Mexico ADR L	21.1	0.14
Matav RT ADR	21.5	0.22
Telstra ADR	21.7	0.12
Gilat Communications	22.7	0.31
Deutsche Telekom AG ADR	24.6	0.11
British Telecommunications PLC ADR	25.7	0.07
Tele Danmark AS ADR	27	0.09
Telekomunikasi Indonesia ADR	28.4	0.32
Cable & Wireless PLC ADR	29.8	0.14
APT Satellite Holdings ADR	31	0.33
Telefonica SA ADR	32.5	0.18
Royal KPN NV ADR	35.7	0.13
Telecom Italia SPA ADR	42.2	0.14
Nippon Telegraph & Telephone ADR	44.3	0.2
France Telecom SA ADR	45.2	0.19
Korea Telecom ADR	71.3	0.44

The Lead In: Data Hypotheses

□ <u>Growth and PE</u>: Companies with higher expected growth should have higher PE ratios. In the table, the companies with lower expected growth rates are trading at lower PE ratios.

□ <u>Risk and PE</u>: Companies that are riskier should trade at lower PE ratios. While these are all telecom firms, some are in developed markets and others are in emerging markets, and in 1997, emerging market companies were viewed as much riskier than developed market companies.

   ◘ I used a dummy variable to capture this, with emerging market companies getting a value of one and developed market companies getting a value of zero.

   ◘ Note that there are richer measures of country risk, like country risk scores and sovereign ratings, which allows for more variation, but they were not available in 1997.

PE, Growth and Risk

- Dependent variable is: PE
- R squared = **66.2%**   R squared (adjusted) = **63.1%**

Variable	Coefficient	SE	t-ratio	Probability
Constant	**13.1151**	3.471	3.78	0.0010
Growth rate	**121.223**	19.27	6.29	≤ 0.0001
Emerging Market Dummy	**-13.853**	3.606	-3.84	0.0009

- You can use this regression to get predicted values for individual stocks. For instance, Indosat has an expected growth rate of 6% and is an emerging market company:

PE = 13.13 + 121.22 (.06) -13.85 (1) = 6.55

At 7.8 times earnings, Indosat is expensive.