

A Review of Cross Sectional Regression for Financial Data

- **You should already know this material from previous study**
- **But I will offer a review, with a focus on issues which arise in finance**

TYPES OF FINANCIAL DATA

Types of Data:

- **time series data (Y_t for $t=1, \dots, T$)**

e.g. stock price every day for several years

- **cross-sectional data (Y_i for $i=1, \dots, N$)**

e.g. data on the share price of $N=100$ companies.

- **Y is referred to as a variable**

(e.g. it varies across companies)

Data Transformations: Levels, Growth Rates, Returns and Excess Returns

- Often original data is transformed

e.g. if X = company earnings, and W = number of shares, you can create a new variable: $Y = X/W$ which is earnings per share.

- When working with asset prices (e.g. price of shares) following transformations are common
- Y_t is the Level = the original variable (e.g. price of share)
- Growth rate is the percentage change

$$\% \text{ change} = \frac{(Y_{t+1} - Y_t)}{Y_t} \times 100$$

- Properties of logarithms can be used to show:
 $\% \text{ change} \approx 100 \times [\ln(Y_t) - \ln(Y_{t-1})]$.

- **Return on an asset is almost the same as its growth rate. But this ignores dividends.**

$$\text{Return} = R_t = \frac{(Y_t - Y_{t-1} + D_t)}{Y_{t-1}} \times 100.$$

- **Excess return is important in much financial decision making**
- **Excess return is the return above what a safe, risk free asset would give:**

$$ER_t = R_t - R_{0t}.$$

R_{0t} is the return on the safe asset (e.g. a government bond)

We will talk more about all these concepts shortly in our CAPM example

Index Numbers

- **Many variables that financial analysts work with come in the form of index numbers.**
- **E.g. The Dow Jones Industrial Average (DJIA) and Standard and Poor's composite share index (S&P500) are stock price indices.**
- **E.g. Suppose you are interested in studying the performance of the stock market as a whole. An index will measure this.**
- **The price of the stock of an individual company (e.g. Microsoft, Ford or Walmart, etc.) can be readily measured. Index measures price of stock market as a whole.**
- **Price index takes the weighted average of share prices of a set of companies (e.g. the biggest 500 companies).**
- **“weights” in weighted average are values of each company (so big companies have more weight than small)**

Index Numbers (continued)

- **Indices are usually normalized to that a base year is 100 (or 10,000 or some base)**
- **E.g. suppose an index is $Y_1=100$, $Y_2=106$, $Y_3=109$ and $Y_4=111$.**
- **First year has been selected as a base year and, accordingly, $Y_1=100$.**
- **Figures for other years are all relative to base year and allow for a simple calculation of how prices have changed since the base year.**
- **For instance, $Y_2=106$ means that prices have risen from 100 to 106 – a 6% rise since the first year. It can also be seen that prices have risen by 9% from year 1 to year 3 and by 11% from year 1 to year 4.**
- **A price index is very good for measuring *changes* in prices over time, but should not be used to talk about the *level* of prices. It should not be interpreted as an indicator of whether prices are “high” or “low”.**

Obtaining Data

- Many web sources (often though, they cost money)
- Popular ones are DataStream (<http://www.datastream.com/>) and Wharton Research Data Services (<http://wrds.wharton.upenn.edu/>).
- For free sources try the Federal Reserve Bank of St Louis (<http://research.stlouisfed.org/fred2/>) or the Financial Data Finder provided by the Fisher College of Business at the Ohio State University
- Many academics also make the data sets they have used available on their websites. For instance, Robert Shiller at Yale University has a good website.

Example: Executive compensation and profits of 70 companies

Question of interest:

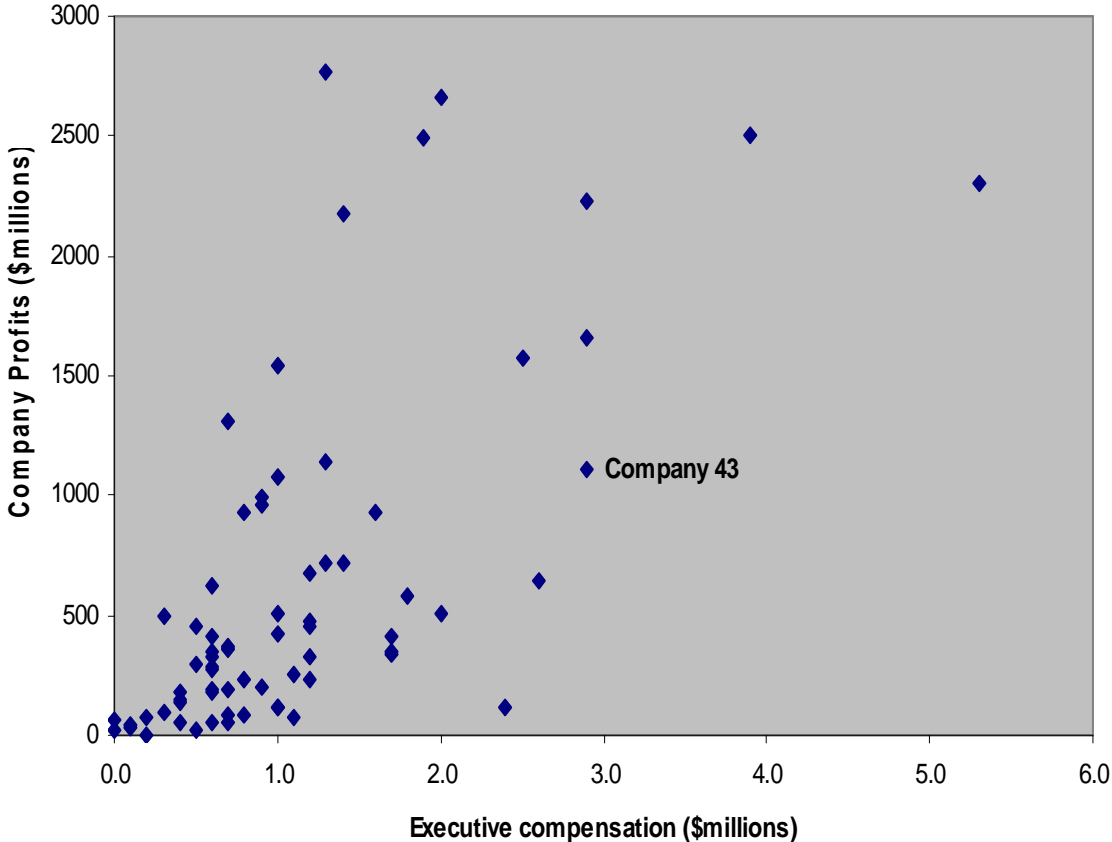
“Do companies with higher profits also pay their executives more?”

Plot of one variable versus another (e.g. profits on y-axis, executive compensation is on x-axis).

Each point on graph represents profits and executive compensation for one company.

See Figure

Figure 2.3: XY-plot of executive compensation against profits



Interpretation of XY-plots

There seems to be a positive relationship between executive compensation and profits

Companies with low profits also tend to have low executive compensation (i.e. low-low)

Companies with high profits also tend to have executive compensation (i.e high-high)

A negative relationship would have low-high and high-low.

Outliers: Companies which do not fit the general pattern.

Descriptive Statistics

Descriptive statistics are *numbers* which summarise properties on income distribution.

1. Measures of location.

Intuition: centre of distribution, average, typical country.

Mean

$$\bar{Y} = \frac{\sum Y_i}{N}$$

Note:

$$\sum = \sum_{i=1}^N$$

Descriptive Statistics (cont.)

2. Measures of dispersion.

Intuition: spread/variability/dispersion of distribution, inequality.

Standard deviation:

$$s = \sqrt{\frac{\sum (Y_i - \bar{Y})^2}{N-1}}$$

Variance = standard deviation squared.

Expected Values and Variances

- **Above we talked about means and variances. Formally these are “sample” means and “sample” variances.**
- **“Sample” emphasizes that they are calculated using an actual “sample” of data.**
- **Population means and population variance are related concepts which we will try and motivate using an example**
- **Suppose we have data on the return to holding stock in a company for the past 100 months. We can use this data to calculate the sample mean and variance.**
- **However, these numbers are calculated based on the historical performance of the company. What about predicting future stock returns?**
- **We do not know exactly what these will be, so cannot calculate sample means and variances as we did above. But a potential investor would be interested in some similar measures.**

- **Investor would be interested in the typical return which she might expect.**
- **She might also be interested in the risk involved in purchasing the stock.**
- **The concept of a typical expected value sounds similar to the ideas we discussed relating to the mean. The concept of riskiness sounds similar to the idea of a variance.**
- **We need concepts like the sample mean and variance, but for cases when we do not actually have data to calculate them. The relevant concepts are the population mean and population variance.**
- **Let Y denote next month's return on this stock. From the investor's point of view, Y is unknown.**
- **The typical return she might expect is measured by the population mean and is referred to as the expected value. We use the notation $E(Y)$ to denote the expected return.**

- **The “expected value” sheds light on what we expect will occur.**
- **However, the return on a stock is rarely exactly what is expected (i.e. rarely will you find Y to turn out to be exactly $E(Y)$).**
- **Stock markets are unpredictable, there is always risk associated with purchasing a stock. Variance relates to this risk. We use the notation $\text{var}(Y)$ for this.**
- **Details about calculate expected values and variances are in the textbook.**

Correlation

Correlation measures numerically the relationship between two variables X and Y (e.g. (e.g. executive compensation and profit).

Correlation between X and Y is symbolised by r or r_{XY} .

Example: Y = executive compensation, X = profit we obtain $r = .66$.

What does the fact r is positive mean?

- **There is a positive relationship between executive compensation and profit.**
- **Companies with high profits tend to pay their executives well.**
- **Companies with low profits tend to have low executive compensation.**
- **Executive compensation and profits both vary across companies. The high/low variation in executive compensation tends to match up with the high/low variation in profits.**

Covariances and Population Correlations

In Chapter 2 distinguished between sample and population means and variances.

Remember sample mean was denoted by \bar{Y} and was the average calculated using the data at hand.

Population mean denoted by $E(Y)$ and called the expected value.

Population variance: $\text{var}(Y)$.

The same sample/population distinction holds with correlations.

Let ρ be population correlation (remember r is the sample correlation).

Discussion and motivation of population correlation given in Chapter 3 along with technical details.

Regression

- **Regression and extensions are the most common tool used in finance.**
- **Used to help understand the relationships between many variables.**
- **Simple Regression Model**

$$Y = \alpha + \beta X + e$$

where e is an error

- **Notation: $\hat{\alpha}$ and $\hat{\beta}$ are the estimates of α and β .**
- **Ordinary least squares (OLS) estimator is the most common one**

Example: Executive Compensation and Profits

Y = executive compensation (millions of \$) = dependent variable

X = profits (millions of \$) = explanatory variable

Using data on N = 70 companies we find:

$$\hat{\beta} = 0.000842$$

Interpretation of $\hat{\beta}$

“If profits increase by \$1 million, then executive compensation will tend to increase by .000842 millions dollars (i.e. \$842)”

R^2 : A Measure of Fit

R^2 measures the proportion of the variability in Y that can be explained by X.

Example:

In regression of Y = executive compensation on X = profits we obtain $R^2=0.44$

“44% of the cross-company variation in executive compensation can be explained by the cross-company variation in profits”

Statistical Aspects of Regression

- **Uncertainty about accuracy of the estimate $\hat{\beta}$ can be summarised in a “confidence interval”**
- **Hypothesis testing is another common activity.**
- **Testing $\beta=0$ is commonly done with t-statistic**
- **Multiple regression same as simple regression except many explanatory variables: X_1, X_2, \dots, X_k**
- **We will illustrate and motivate regression with the following extended financial example**

An Extended Finance Regression Example: The Capital Asset Pricing Model

- **The Capital asset pricing model (CAPM) is a very popular model and it, or extensions of it are widely used by financial analysts.**
- **We show how it yields a simple regression which can be estimated using OLS.**
- **A key concept in finance: If we let Y_t be the price of a stock and D_t be the dividend paid between period t-1 and t then the *return* made by an investor (measured as a percentage) would be**

$$\text{Return} = R_t = \frac{(Y_t - Y_{t-1} + D_t)}{Y_{t-1}} \times 100.$$

- **Another key concept: the *excess return*.** This is the difference between an asset's return and that of a safe, low-risk asset (e.g. the return on a government bond). If the return on a risk-free asset is denoted by R_{0t} , then the excess return (ER), measured as a percentage, on an asset is defined as:

$$ER_t = R_t - R_{0t}.$$

- **An important issue when deciding on an investment portfolio is the tradeoff between risk and expected return.**
- **Some assets are very safe (e.g. government bonds or cash), while others are moderately safe (e.g. purchasing the shares of an established Blue Chip company) and others are much more risky (e.g. purchasing the shares of a newly established Dotcom).**
- **There is a large theoretical literature on *mean-variance efficient portfolios* that trade off risk and return in an optimal manner. The CAPM is an implication of this theory.**

- To explain the CAPM, we need to define the returns on three types of assets.
- R is the return on the asset under study (e.g. the return on holding a share in a particular company, call it Company A).
- R_f is the risk free rate of interest (e.g. the return on a very safe investment such as a government bond).
- R_m be the return on the market portfolio. The market portfolio is formally defined as a portfolio containing the shares of every possible company with portfolio weights proportional to each company's market capitalization. In empirical work, the market portfolio is usually proxied using a stock market index such as the S&P500, NYSE or FTSE.
- We will not derive the CAPM, but merely state some of its main implications and provide some intuitive motivation for it.

- **When deciding on a portfolio, the investor does not know for certain the return of the stocks that are available, hence we have to talk in terms of expected returns (use “E” as notation for “expected”). It is assumed that the return on the risk free asset is known.**

- **The CAPM implies:**

$$E(R) = R_f + \beta[E(R_m) - R_f]$$

- **In words, the expected return on the share of Company A is equal to the risk free rate plus β times the expected excess return on the market portfolio.**
- **Remember that an excess return is a return minus the return on a safe asset.**

- **Intuition:** Investors could have bought the safe asset or the market portfolio. Investors' decisions on whether to buy shares in Company A will thus depend on the returns on the other available options. And investors' decisions determine the price of shares in Company A. Thus, its expected return can reasonably be expected to depend on R_f and $E(R_m)$.
- The precise relationship depends on β , which is commonly referred to as the CAPM β or investor's β .
- The CAPM β depends on the correlation between the return on Company A's shares and the return on the market portfolio.
- **Intuition:** it is wise for an investor to diversify portfolio.

More Intuition: Case 1: Company A Matches the Stock Market

- **Consider what happens when the Company A's shares are just as volatile as the stock market as a whole and movements in the one perfectly match the other.**
- **Every time the stock market rises, Company A's shares rise by the same amount. Every time the stock market falls, Company A's shares fall by the same amount.**
- **In Case 1, CAPM theory implies $\beta=1$ and the CAPM equation implies that both expected returns have to be the same as one another.**
- **If shares in Company A and the stock market as a whole are exactly as risky as one another then their expected returns must be the same. If the stock market index had a higher expected return, then no investor would ever buy the (equally risky, but with lower expected return) shares of Company A.**

More Intuition: Case 2: Company A Negatively Correlated with the Stock Market

- **In contrast, consider what would happen if returns to Company A's stock were negatively correlated with the stock market as a whole. Then, CAPM theory implies that β is negative.**
- **Note that the expected excess return on the market portfolio is positive (since if $E(R_m) < R_f$ then no one would ever invest in the stock market – they would prefer the safe higher return of the risk free asset).**
- **Using these facts in the CAPM equation, it can be seen that they imply $E(R) < R_f$ and the excess return to Company's A shares is actually negative!**
- **How can this be? In Case 2, Company A's shares would be very attractive to investors. After all, if their returns are negatively correlated with the stock market as a whole, then whenever there is a stock market crash, the shares in Company A would actually rise.**

- **In fact, in a sense they are even safer than a risk free asset such as cash. In a stock market crash, cash would still hold its value, but shares in Company A would actually increase in value.**
- **To an investor who hates risk, they would be an ideal purchase. Such investors would be willing to buy shares in Company A even with their expected return was lower than the risk free rate.**

- **The Case 1 and Case 2 stories are intended to motivate the CAPM equation and its importance for investment decisions.**
- **In practice, the CAPM β can take a wide range of values and it is crucial for making deciding an investment portfolio.**
- **Given its importance for investment decisions, financial analysts often try to estimate what the CAPM β is for individual company's shares.**
- **This can be done using regression methods.**
- **One cannot obtain data on the expected excess return of a share or a market portfolio. However, if we replace expected returns with the actual returns of a stock from the past we can run a regression based on the CAPM equation.**
- **That is, we set $Y =$ excess returns on Company A's shares (measured as the actual return minus the return on a risk free asset such as a government bond).**

- **X = excess returns on the market portfolio (measured as the actual return on a major stock market index minus the return on a risk free asset such as a government bond)**
- **The file CAPM.XLS contains monthly data on Y = the excess returns on Company A's stock for the last 10 years as well as data on X = excess returns on a stock market index for the same time period.**
- **If we run the regression of Y on X we obtain $\hat{\alpha} = 0.43$ and $\hat{\beta} = 1.77$. Thus, our estimate of the CAPM β for Company A is 1.77.**
- **The 95% confidence interval corresponding to our estimate is [1.37,2.17]. Thus, we can conclude that, with high confidence, the CAPM β is at least 1.37 and at most 2.17.**
- **With the CAPM, the hypothesis that $\beta=0$ is rarely tested. It implies that the excess return on the stock of a particular company is completely unrelated to the market portfolio (something which rarely occurs in practice).**

- **However, the original CAPM equation did not contain an intercept. Thus, the CAPM implies that there should be no intercept in the regression of $Y =$ the excess returns on Company A's stock on $X =$ excess returns on a stock market index.**
- **This motivates an interest in testing the hypothesis that $\alpha=0$. The hypothesis test of $\alpha=0$ can be done in exactly the same manner as testing $\beta=0$. That is, you can look at the P-value corresponding to the intercept and check whether it is less than .05.**
- **With our data set we find that the P-value is .052. Thus, it is (very slightly) larger than .05 and we can (very marginally) accept the hypothesis that $\alpha=0$ at the 5% level of significance.**
- **But we would reject the hypothesis that $\alpha=0$ at 10% level of significance.**

A Multiple Regression using CAPM

The data set CAPM.XLS, contains the following variables:

- Y = the excess return on shares in Company A (a percentage)
- X_1 = the excess return on a stock index (a percentage)
- X_2 = the sales of Company A (thousands of dollars)
- X_3 = the debt of Company A (thousands of dollars).

- The theory underlying the CAPM implies that explanatory variables sales and debt should not be significant.

Results from OLS Regression

	Coefficient	Stand Error	t Stat	P-value	Lower 95%	Upper 95%
Inter.	2.530	1.335	1.895	0.061	-0.115	5.175
X1	1.747	0.202	8.631	4E-14	1.346	2.148
X2	-0.0003	0.001	-0.323	0.748	-0.002	0.002
X3	-0.022	0.012	-1.867	0.064	-0.045	0.001

Furthermore, $R^2=.41$ and the P-value for testing $R^2=0$ is $2.83E-13$.

- If we use the 5% level of significance we do find that these two explanatory variables are not significant.
- However, since the P-value for testing whether $\beta_3=0$ is less than 0.10, the debt variable is significant at the 10% level.
- This illustrates how the choice of significance level can have an important effect on the outcome of a hypothesis test.