Prof. Lorenzo Naranjo Fall 2025

The Index Model

Introduction

One of the disadvantages of the Markowitz model is that it requires estimates of all covariances and expected returns. Estimating the covariance between two assets is not hard, and can be done with precision, but estimating the pairwise covariances of five hundred different assets is a large number! Indeed, for five hundred stocks we would have to estimate $500 \times 501/2 = 125{,}250$ different covariances. One way to circumvent this problem is to realize that stock returns exhibit a factor structure driven by the market.

Indeed, for many stocks the market explains a large fraction of their variability. Whatever is not explained by the market is firm-specific risk. The index model formalizes this intuition by splitting the variance of each stock into systematic and idiosyncratic variance.

The Model

The single index model is a linear regression between the excess returns of a stock and the excess returns of the market portfolio. Let's denote by r_i the return of stock i over a given period, and define $R_i = r_i - rf$ as the excess return over the risk-free asset. Similarly, denote by R_m the excess return of the market over the risk-free rate. The index model postulates

$$R_i = \alpha_i + \beta_i R_m + e_i, \tag{1}$$

where $\mathrm{E}(e_i) = 0$ and $\mathrm{Cov}(R_m, e_i) = 0$.

In the single index model, the beta with the market captures the exposure of all securities to systematic risk, which is the risk shared by all securities. Therefore, the error term is idiosyncratic. The single index model assumes that $\mathrm{Cov}(e_i, e_j) = 0$ for any two securities i and j whose returns are not perfectly correlated.

Variance Decomposition

The assumptions of the single index model imply that the variance of R_i can be split into two parts:

$$\begin{aligned}\sigma_i^2 &= \mathrm{V}(\alpha_i + \beta_i R_m + e_i) \\ &= \beta_i^2\,\mathrm{V}(R_M) + \mathrm{V}(e_i) + 2\beta_m\,\mathrm{Cov}(R_m, e_i) \\ &= \beta_i^2\sigma_M^2 + \sigma^2(e_i).\end{aligned} \tag{2}$$

The first component of σ_i^2 is the systematic variance, which depends on the beta of the security but also the variance of the market. The second term of the variance of i is typically computed as the difference between the variance of i and the systematic variance.

Example 1. Suppose that you have the following regression for stock A:

$$R_A = \alpha_A + \beta_A R_M + e_A,$$

where $\alpha_A = 0.02$, $\beta_A = 1.2$, $\sigma(e_A) = 30\%$ and $\sigma_M = 25\%$. The variance of A can be computed as follows:

$$\sigma_A^2 = 1.2^2 \times 0.25^2 + 0.30^2 = 0.09 + 0.09 = 0.18.$$

In the previous expression, the systematic and idiosyncratic variances are the same. The standard deviation of A is then $\sigma_A = \sqrt{0.18} = 42.43\%$. □

The Security Characteristic Line (SCL)

The beta of the model can be estimated from the covariance of R_i and R_m, and the variance of R_m. Indeed, we have that

$$\begin{aligned}\mathrm{Cov}(R_i, R_m) &= \mathrm{Cov}(\beta_i R_m, R_m) + \mathrm{Cov}(e_i, R_m) \\ &= \beta_i\,\mathrm{V}(R_m),\end{aligned}$$

where we use the fact that $\mathrm{Cov}(R_m, R_m) = \mathrm{V}(R_m)$. Thus, in the index model we must have that

$$\beta_i = \frac{\mathrm{Cov}(R_i, R_m)}{\mathrm{V}(R_m)}. \tag{3}$$

The alpha of the security can then be computed as

$$\alpha_i = \mathrm{E}(R_i) - \beta_i \, \mathrm{E}(R_M). \tag{4}$$

The line

$$y = \alpha_i + \beta_i x$$

is called the *security characteristic line* (SCL) of security i. If we plot this line using a line chart, α_i is the intercept and β_i is the slope coefficient of the line.

R-Squared

Following the statistic literature, the proportion of systematic variance to total variance is called the R-squared of security i and can be expressed as

$$\text{R-squared} = \frac{\beta_i^2 \sigma_M^2}{\sigma_i^2} = 1 - \frac{\sigma^2(e_i)}{\sigma_i^2}. \tag{5}$$

Therefore, the R-squared can also be expressed as one minus the proportion of idiosyncratic variance to total variance. Since the single index model aims to decompose the total variance of a security into two orthogonal components, the R-squared gives us the proportions of this decomposition.

Example 2. You regress the excess returns of stock B on the excess returns of the market:

$$R_B = \alpha_B + \beta_B R_M + e_B.$$

Your regression package reports that $\alpha_B = -0.01$, $\beta_B = 0.8$ and the R-squared is 0.4. If the volatility of the market is 25% per year, the systematic variance is $0.8^2 \times 0.25^2 = 0.04$.

Since 40% of the variance is systematic, we have that

$$\sigma_B^2 = \frac{0.04}{0.4} = 0.10,$$

which implies that $\sigma_B = \sqrt{0.10} = 31.62\%$ per year. We also know that 60% of the variance is firm-specific, which means that

$$\sigma^2(e_B) = 0.6 \times 0.10 = 0.06,$$

so that $\sigma(e_B) = \sqrt{0.06} = 24.49\%$ per year. □

Equation (5) can also be expressed in terms of the correlation between R_i and R_M. Indeed, since

$$\beta_i = \frac{\text{Cov}(R_i, R_M)}{\text{V}(R_M)} = \frac{\sigma_i \sigma_M \rho_{i,M}}{\sigma_M^2} = \frac{\sigma_i \rho_{i,M}}{\sigma_M}, \tag{6}$$

we have that

$$\text{R-squared} = \frac{\beta_i^2 \sigma_M^2}{\sigma_i^2} = \frac{\frac{\sigma_i^2 \rho_{i,M}^2}{\sigma_M^2}\sigma_M^2}{\sigma_i^2} = \rho_{i,M}^2. \tag{7}$$

Thus, the R-squared of a regression of R_i on R_M is just the square of the correlation between R_i and R_M. The name R-squared comes from the fact that we typically use the greek letter *rho* (ρ), which corresponds to the latin letter *r*, for correlation.

An Example with Real Data

As an example, let's analyze the monthly returns of Microsoft (Ticker: MSFT) from June 2015 until June 2025. All data comes from Yahoo Finance. As a proxy for the risk-free rate, I use the 13-week Treasury Bill CBOE Index (Ticker ^IRX). The rate is expressed per year, so I convert it to a monthly rate by:

$$r_{\text{monthly}} = (1 + r_{\text{annual}})^{1/12} - 1.$$

To proxy for the market, I use the SPDR S&P 500 ETF Trust (Ticker: SPY), which allows me to include the dividend distribution of the stocks forming the S&P 500. The monthly

returns are computed using the *adjusted price* series to obtain a holding period return that includes dividends.

The table below presents some descriptive statistics of both series of excess monthly returns.

Table 1: The table presents descriptive statistics of Microsoft and S&P 500 monthly returns using data from June 2015 until June 2025.

Ticker	MSFT	SPY
Mean (%)	2.175	1.002
St. Dev. (%)	6.049	4.465

The table shows that during the period, the monthly returns of Microsoft are more volatile than those of the S&P 500, but so is the average excess returns of the two series.

The figure below presents a scatter plot of the data. Clearly, the points cluster around the SCL, and we can see that the range of returns of Microsoft is significantly wider than the S&P 500.

We can use *ordinary least squares* (OLS) to estimate the slope coefficient and the intercept of the SCL. Many statistical packages allow to do this. The results below are computed using the Python library *statsmodels*.

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                   MSFT   R-squared:                       0.522
Model:                            OLS   Adj. R-squared:                  0.517
No. Observations:                 120   F-statistic:                     114.5
Covariance Type:                  HC1   Prob (F-statistic):           4.32e-
19
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      0.0120      0.004      3.195      0.001       0.005       0.019
SPY            0.9784      0.091     10.701      0.000       0.799       1.158
==============================================================================
```

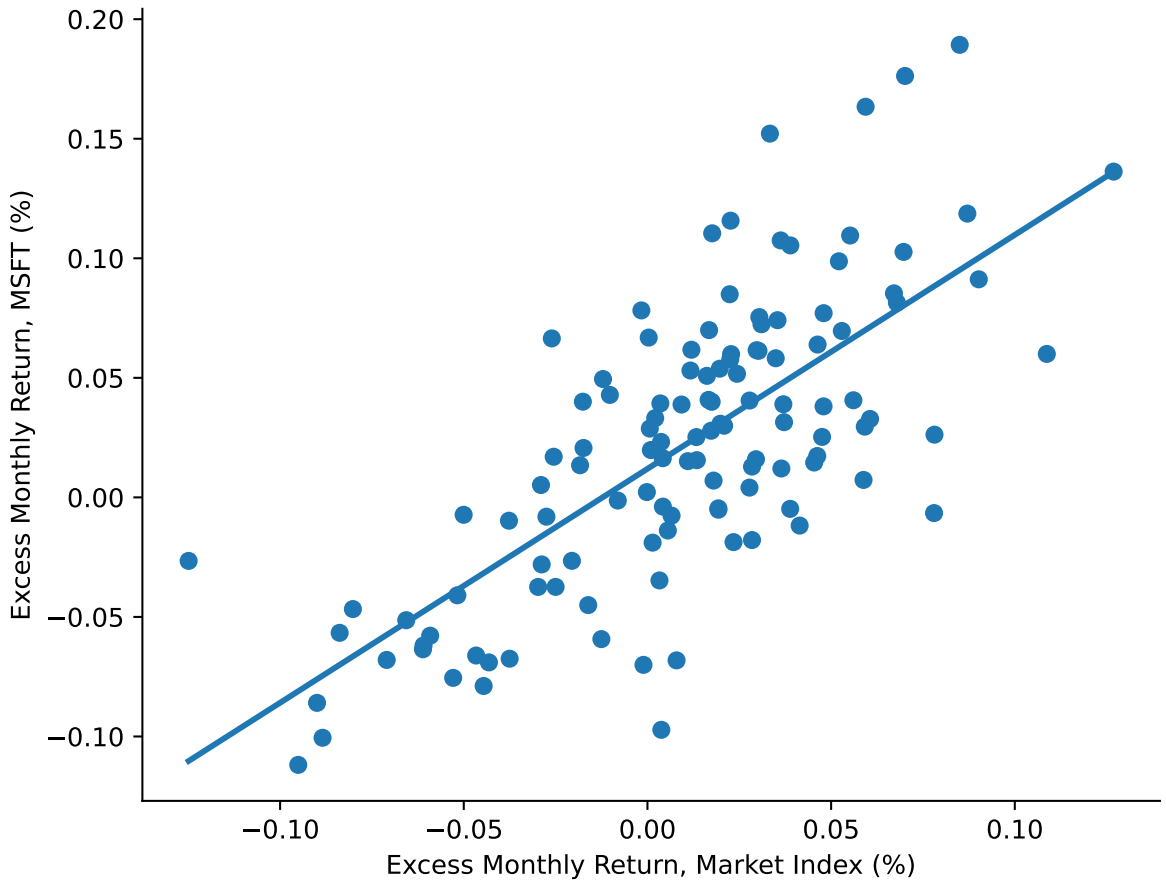

Figure 1: The figure plots the excess monthly returns of the market portfolio, proxied by SPY vs. the excess monthly returns of MSFT from June-2015 until June-2025.

```
Notes:
[1] Standard Errors are heteroscedasticity robust (HC1)
```

The table presents many numbers but for our purposes we can focus on just a few of them. First, under the column `coef` we can see that the estimate for the intercept is 0.0123 whereas the beta estimate for MSFT is 0.9894. Since the beta of Microsoft is close to one, we can conclude that Microsoft carries almost the same systematic risk as the market.

We can compute the R-squared of the regression using the beta and variance of MSFT, and the variance of the market:

$$\text{R-squared} = \frac{0.9894^2 \times 4.405^2}{6.239^2} = 0.488.$$

The variance of MSFT and the S&P 500 are computed by squaring the standard deviations reported in Table 1. We can see that the computed R-squared corresponds to the R-squared reported by the regression package, implying that 48.9% of the variance is explained by the exposure of MSFT to the market. The remaining variance is firm-specific risk.

Differences in Beta and R-Squared

The figure below shows a scatter plot of market excess monthly returns vs. the monthly excess returns of two financial and two technology stocks:

- Citigroup (Ticker: C)
- BlackRock (Ticker: BLK)
- Nvidia (Ticker: NVDA)
- Tesla (Ticker: TSLA)

As before, we proxy the the market portfolio using the SPDR S&P 500 ETF Trust (Ticker: SPY), which allows us to include the dividend distribution of the stocks forming the S&P 500. The risk-free rate is obtained from the 13-week Treasury Bill CBOE Index (Ticker ^IRX).

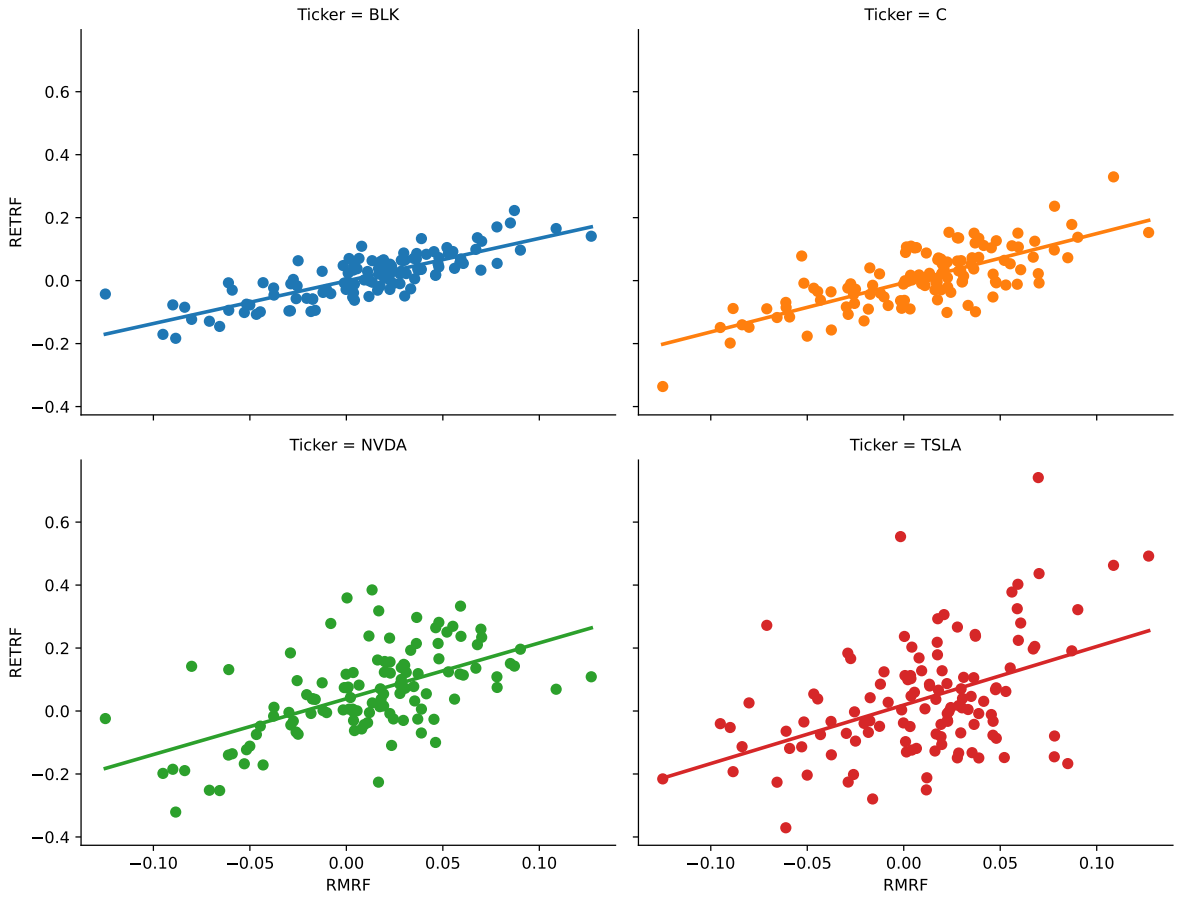

Figure 2: The figure plots the excess monthly returns of the market portfolio labeled as RMRF and proxied by SPY, vs. the excess monthly returns of Citigroup (C), BlackRock (BLK), Nvidia (NVDA), and Tesla (TSLA) labeled RETRF, from June-2015 until June-2025.

The picture shows that different stocks have different degrees of firm-specific risk. BlackRock excess returns align quite well with market returns, whereas Tesla excess returns are the most dispersed. In general, you would expect more dispersion in technology stocks since by definition new technologies might or might not work. The fact that you get a technology to work is independent of what the market does.

The table below present relevant values of the regression.

Table 2: The table presents the Alpha, Beta and R-Squared estimates obtained by running a linear regression of excess monthly returns for Citigroup (C), BlackRock (BLK), Nvidia (NVDA), and Tesla (TSLA) on the market portfolio proxied by SPY using data from June-2015 until June-2025.

	Alpha		Beta	R-Squared
	Estimate	P-value		
BLK	-0.0011	0.804	1.353	0.668
C	-0.0068	0.239	1.562	0.550
NVDA	0.0390	0.000	1.772	0.337
TSLA	0.0190	0.182	1.859	0.206

First, we can see that the R-squared is the lowest for Tesla, which is apparent from the pictures. The market explains a small fraction of the variance for the stock. On the other hand, almost 67% of BlackRock's variance is explained by the market.

Also, we see that both Nvidia and Tesla load on significant systematic risk. Certainly, their cash flows are exposed to how the economy does and this is reflected on their high betas. Citigroup also has a high beta, which is typical of financial firms that are also exposed to how the market performs. BlackRock has the lowest beta of the four stocks, although is still higher than the beta we estimated for Microsoft.

During the period the only stock that has a positive alpha statistically different from zero at the 1% significance level is Nvidia. For the other stocks, we cannot reject the null hypothesis that the alpha is different from zero. The alpha of Nvidia is indeed impressive. It has out-performed the market by 3.90% per month during the last 10 years.

Finally, the table below shows the pairwise correlations between each stock and the market.

Table 3: The table shows the pairwise correlations and the square of their values between excess monthly returns for Citigroup (C), BlackRock (BLK), Nvidia (NVDA), and Tesla (TSLA) and the market portfolio proxied by SPY using data from June-2015 until June-2025.

	Correlation	Correlation Squared
BLK	0.818	0.668
C	0.741	0.550
NVDA	0.580	0.337
TSLA	0.454	0.206

As expected, the square of the correlation corresponds to the R-squared reported in Table 2.

Implications of the Model

When you run a regression of the excess returns of a security on the excess returns of the market, the residuals are automatically orthogonal to the regressor. Therefore, in equation (1) we must have that $\text{Cov}(R_m, e) = 0$ for all securities.

The crucial assumption of the single index model is that the only systematic source of risk is the exposure of each security to the market. The implication of this assumption is that the covariance of the residuals between two securities is zero as long as their excess returns are not perfectly correlated.

We saw before that if A and B are perfectly correlated, we must have that $R_A = wR_B$ for some $w \neq 0$. If this was the case, the covariance between e_A and e_B is not zero even though they are different assets. More precisely, we have that

$$\text{Cov}(e_A, e_P) = \text{Cov}(e_A, we_A) = w\sigma^2(e_A) \neq 0,$$

provided that $w \neq 0$, i.e., you do not invest everything in the risk-free asset.

IN the following, when we talk about two different assets it is implicitly assumed that their excess returns are not perfectly correlated, unless stated otherwise.

Covariance Structure

The covariance of the excess returns between securities i and j is

$$\begin{aligned} \text{Cov}(R_i, R_j) &= \text{Cov}(\alpha_i + \beta_i R_M + e_i, \alpha_j + \beta_j R_M + e_j) \\ &= \beta_i \beta_j \, \text{Cov}(R_M, R_M) \\ &= \beta_i \beta_j \sigma_M^2. \end{aligned} \tag{8}$$

The previous expression says that in the single index model, the covariance between of any two different securities is given by their exposures to the market and the variance of the market. We can express equation (8) in terms of correlations as

$$\begin{aligned} \rho_{i,j} &= \frac{\text{Cov}(R_i, R_j)}{\sigma_i \sigma_j} = \frac{\beta_i \beta_j \sigma_M^2}{\sigma_i \sigma_j} \\ &= \frac{\frac{\sigma_i \rho_{i,M}}{\sigma_M} \frac{\sigma_j \rho_{j,M}}{\sigma_M} \sigma_M^2}{\sigma_i \sigma_j} \\ &= \rho_{i,M} \rho_{j,M}. \end{aligned} \tag{9}$$

Therefore, in the single-index model, all pairwise correlations between two assets can be computed as the product of their correlations with the market.