Correlate Pro

1 Notes

1.1 Intro

In Correlate Pro, we represent a portfolio of N stocks as an undirected graph G where G_{ij} is the correlation coefficient between stocks i and j. Sometimes, we view the entire, fully-connected graph while at other times we fixate on a stock, x, and only show the correlations between x and every other stock.

1.2 Motivation

Why do we care about correlation coefficient? Well, conventional wisdom dictates that our investments should be diverse, i.e. we don't want all of our eggs in one basket. Consequently, investors tend to pick companies in different industries to invest in. For example, a sophisticated investor might choose to invest in real estate, tech, auto, finance, and the energy industry. In some ways, each of these industries impact each other but we wouldn't say that Apple is a competitor to Goldman Sachs or Marathon. Therefore, if one of these industries falls the investor's other industries should be unaffected and weather the storm.

However, the lines between industries are blurring. In today's economy, many companies across different industries might be owned by one parent company or one company may have products spanning multiple sectors. One of the most notable examples is Tesla. Is it a car company? Tech company? Energy company? All of the above. Because companies like Tesla are spreading business across multiple industries, we have to look at the math in order to tell if our portfolio is diverse. A rule of thumb oversimplifies the modern economy.

1.3 Math

Say we have two stocks, x and y, whose returns over the past year are stored in vectors $\vec{x}$ and $\vec{y}$. Let's also say the length of these vectors is n (if we have daily closes over the past year, then $n = 252$). We want to figure out how similar the patterns are in their returns.

Now, we only care about the proportions, e.g. if x goes up 5% does y also go up 5%? Not if x goes up \$5 does y go up \$5. To account for this, we first normalize the returns:

$$\vec{x}_s = \frac{\vec{x} - \mu_x}{\sigma_x}$$

$$\vec{y}_s = \frac{\vec{y} - \mu_y}{\sigma_y}$$

Where μ_x, μ_y are the averages for x and y's returns, respectively, and σ_x, σ_y are the standard deviations for x and y's returns. To get the correlation coefficient, we use $\vec{x}_s$ and $\vec{y}_s$:

$$G_{ij} = \frac{1}{n-1}\vec{x}_s \cdot \vec{y}_s$$

Below are some examples of $\vec{x}$ vs $\vec{y}_s$ and the consequent correlation coefficients:

$r = -1$

(a)

$r = -0.94$

(b)

$r = +0.08$

(c)

$r = +1$

(d)

$r = +0.86$

(e)

$r = 0.00$

(f)