Correlation Measures

Pedro Giraldi

Covariance

- Discrete variables: $Cov(x, y) = \sigma_{12} = \frac{\sum_i (x_{i1} - \overline{x_1})(x_{i2} - \overline{x_2})}{N}$

- Continuous variables:

$$Cov(x, y) = \sigma_{12} = \int_1 \int_2 [x_1 - E(X_1)][x_2 - E(X_2)]\, f_{12}(x_1, x_2)\, dx_2\, dx_1$$

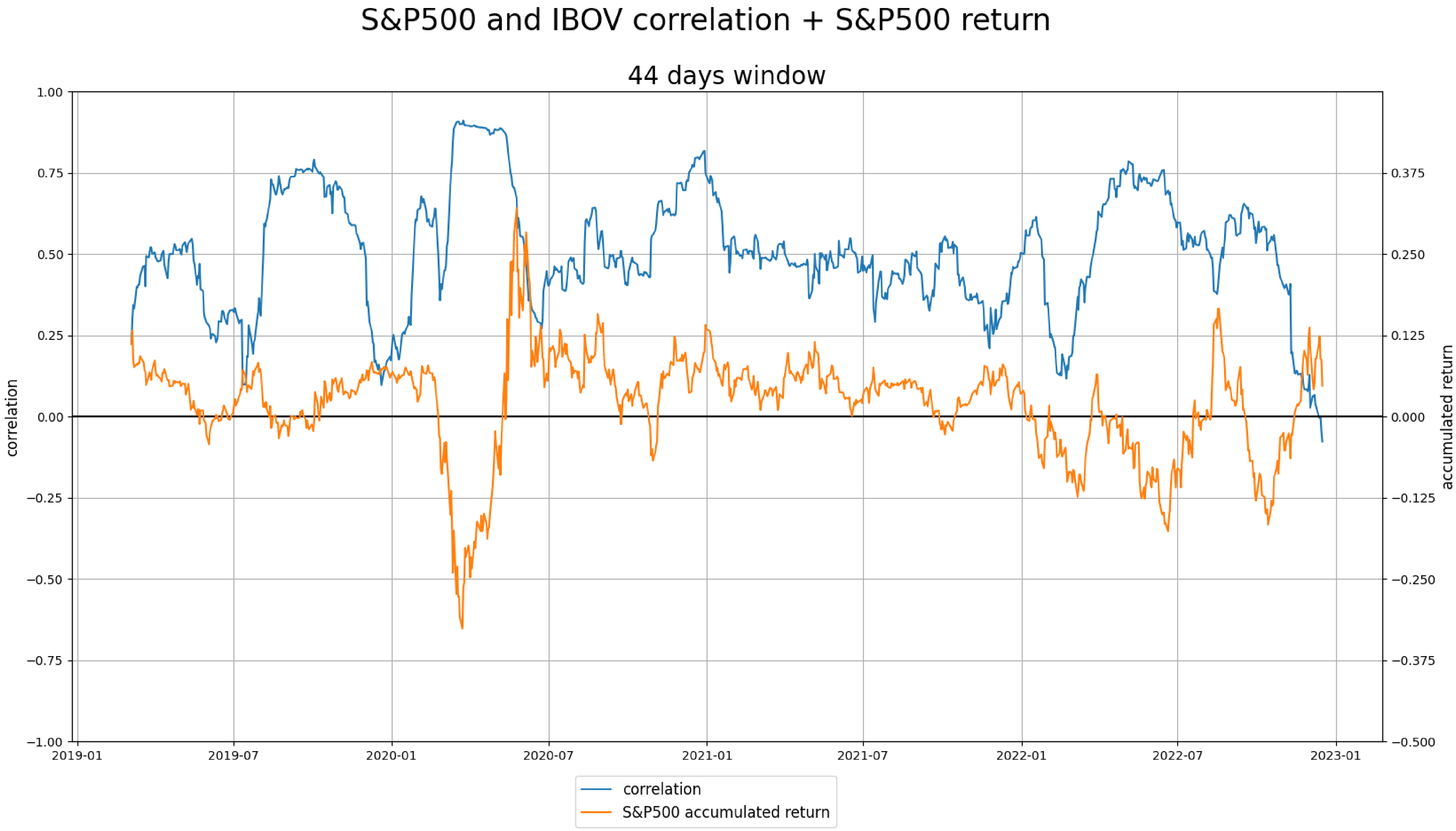
$given$: $$f_{12}(x_1, x_2) = f_1(x_1) * f_2\,(x_2) * c_{12}$$

Covariance between a discrete random variable and itself

$$Cov(x,x) = \sigma_{11} = \frac{\sum_i (x_{i1} - \overline{x_1})(x_{i1} - \overline{x_1})}{N} = \frac{\sum_i (x_{i1} - \overline{x_1})^2}{N} = \sigma_1^2$$

Correlation: normalized covariance:

$$\rho_{12} = \frac{Cov(x_1, x_2)}{\sigma_1 \sigma_2} \mid \rho \in [-1,1]$$


S&P500 and IBOV correlation + S&P500 return
44 days window
correlation
accumulated return
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00
0.375
0.250
0.125
0.000
−0.125
−0.250
−0.375
−0.500
2019-01
2019-07
2020-01
2020-07
2021-01
2021-07
2022-01
2022-07
2023-01
correlation
S&P500 accumulated return

Defining a portfolio variance formula

$$W = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix}, COV = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_n^2 \end{bmatrix}$$

$$\sigma_p^2 = WT * COV * W$$

For a portfolio with two assets:

$$\sigma_p^2 = \begin{bmatrix} w_1 & w_2 \end{bmatrix} * \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} * \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$$

$$\sigma_p^2 = w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_{12}$$

Analyzing the portfolio variance formula

$$\sigma_p = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{12}\sigma_1\sigma_2}$$

Let $\rho_{12} = 1$:

$$\sigma_p = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_1\sigma_2}$$

$$\sigma_p = \sqrt{(w_1\sigma_1 + w_2\sigma_2)^2} = w_1\sigma_1 + w_2\sigma_2$$

Analyzing the portfolio variance formula

$$\sigma_p = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{12}\sigma_1\sigma_2}$$

Now let $\rho_{12} < 1$:

$$\sigma_p = \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\rho_{12}\sigma_1\sigma_2} < \sqrt{w_1^2\sigma_1^2 + w_2^2\sigma_2^2 + 2w_1w_2\sigma_1\sigma_2}$$

$$\sigma_p < w_1\sigma_1 + w_2\sigma_2$$

Effect on portfolio variance of increasing the number of assets

$$\sigma_p^2 = \sum_{i=1}^{N} w_i^2 \sigma_i^2 + \sum_{i,\, j,\, i \neq j}^{N} w_i w_j Cov(i,j) = \sum_{i,\, j}^{N} w_i w_j Cov(i,j)$$

For large N, correlation becomes the main determinant of portfolio variance

$$\sigma_p^2 = \frac{\bar{\sigma}^2}{N} + \frac{(N-1)}{N}\overline{Cov} = \frac{\bar{\sigma}^2}{N} + \frac{(N-1)}{N}\bar{\rho}\bar{\sigma}^2 \cong \bar{\rho}\bar{\sigma}^2$$

Beta

Beta is a measure of an asset's sensitiveness to a benchmark, and corresponds to the slope of the regression between returns.

There are different ways to estimate it, but all derive from the same method:

1. Perform a regression between the benchmark (independent, x) and the asset (y)
2. Use the unbiased OLS beta estimator
3. Use an alternative version of the OLS beta estimator

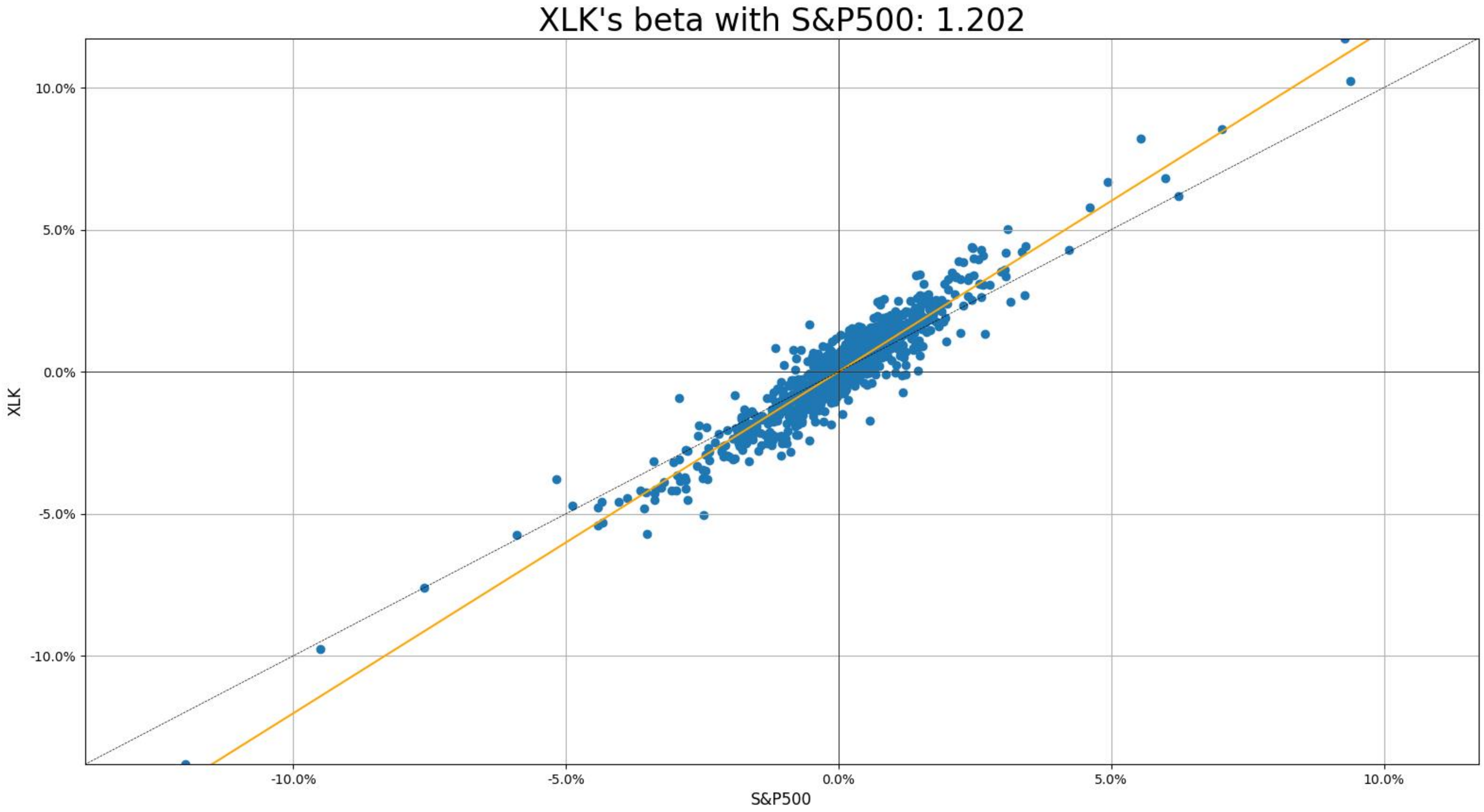
#1: regression

There are many Python packages to calculate regressions, but statsmodels is the most intuitive.

Using the .summary() method, the regression's R-squared and the Beta's p-value help checking for statistical significance

```
                                 OLS Regression Results
=======================================================================================
Dep. Variable:                    XLK   R-squared (uncentered):                   0.894
Model:                            OLS   Adj. R-squared (uncentered):              0.894
Method:                 Least Squares   F-statistic:                              8394.
Date:                Tue, 20 Dec 2022   Prob (F-statistic):                        0.00
Time:                        15:14:27   Log-Likelihood:                          3697.8
No. Observations:                1000   AIC:                                     -7394.
Df Residuals:                     999   BIC:                                     -7389.
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
S&P500         1.2022      0.013     91.621      0.000       1.176       1.228
==============================================================================
Omnibus:                       36.027   Durbin-Watson:                   1.970
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               95.384
Skew:                           0.028   Prob(JB):                     1.94e-21
Kurtosis:                       4.512   Cond. No.                         1.00
==============================================================================

Notes:
[1] R² is computed without centering (uncentered) since the model does not contain a constant.
[2] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```



XLK's beta with S&P500: 1.202
10.0%
5.0%
0.0%
-5.0%
-10.0%
XLK
-10.0%
-5.0%
0.0%
5.0%
10.0%
S&P500

#2: unbiased Beta estimator

$$\beta_{i,m} = \frac{cov(R_i, R_m)}{\sigma_m^2}$$

Given:

R_i: asset returns

R_m: market/benchmark returns

σ_m^2: market/benchmark variance

#3: modifying the unbiased Beta estimator

$$\beta_{i,m} = \frac{cov(R_i, R_m)}{\sigma_m^2} = \frac{\rho_{i,m}\sigma_i\sigma_m}{\sigma_m^2} = \frac{\rho_{i,m}\sigma_i}{\sigma_m}$$

Note that this formula divides the Beta coefficient in two parts: the correlation between the asset and the market and their volatility ratio.

Converting a Covariance Matrix into a Beta Matrix

Given:

$$COV = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_n^2 \end{bmatrix}$$

Define A:

$$A = \begin{bmatrix} \frac{1}{\sigma_1^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sigma_n^2} \end{bmatrix}$$

Then:

$$COV * A = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_n^2 \end{bmatrix} * \begin{bmatrix} \frac{1}{\sigma_1^2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sigma_n^2} \end{bmatrix} = \begin{bmatrix} \frac{\sigma_1^2}{\sigma_1^2} & \cdots & \frac{\sigma_{1n}}{\sigma_n^2} \\ \vdots & \ddots & \vdots \\ \frac{\sigma_{n1}}{\sigma_1^2} & \cdots & \frac{\sigma_n^2}{\sigma_n^2} \end{bmatrix} = \begin{bmatrix} 1 & \cdots & \beta_{1,n} \\ \vdots & \ddots & \vdots \\ \beta_{n,1} & \cdots & 1 \end{bmatrix}$$

Interpreting a Beta Matrix

$\downarrow$ Assets / Benchmarks $\rightarrow$	XLK	S&P 500
XLK	1	1.201
S&P 500	0.744	1

The Beta coefficient of XLK in relation to the S&P 500 is 1.2, which means for each 1bp. move in the S&P, we can expect a 1.2bp. move in in the XLK in the same direction.

Note that $\beta_{1,2} \neq \frac{1}{\beta_{2,1}}$, why?

Condition for β Invertibility

$$\beta_{x,y} = \frac{1}{\beta_{y,x}} \rightarrow \frac{cov(x,y)}{\sigma_y^2} = \frac{\sigma_x^2}{cov(x,y)}$$

$$cov^2(x,y) = \sigma_x^2 \sigma_y^2$$

$$\frac{cov^2(x,y)}{\sigma_x^2 \sigma_y^2} = 1$$

$$\rho_{1,2}^2 = 1$$

$$\rho_{1,2} = \pm 1$$