Econometrics

1st **Lab**

Élèves :
Thomas DE PORTZAMPARC

Enseignants :
Gaëlle LE FOL
Prénom NOM

19 septembre 2025

Table des matières

Introduction

This project explores the relationship between the **CAC40 index** and other major financial indices, namely the **EuroStoxx50**, the **Dow Jones**, and the **S&P500**. Using Python, we collected and processed daily return data from multiple sources, stored in both a combined Excel file and separate CSVs. The goal was to determine whether movements in the CAC40 can be explained by or linked to the dynamics of other indices.

The work involved several key steps :

- **Data preparation** : checking formats, cleaning, aligning series, and handling missing values.
- **Return calculations** : daily and monthly returns were computed to capture short- and medium-term dynamics.
- **Statistical analysis** : descriptive statistics, covariance, and correlation structures were analyzed.
- **Regression modeling** : linear regressions were used to quantify dependencies, with tests for intercept and slope significance.
- **Visualization** : scatter plots, time series graphs, and regression plots provided an intuitive view of the results.

Overall, the project combines **financial time series analysis and econometric modeling** to assess cross-market relationships and dependencies.

Expected Results

Given the definition and composition of the indices under study, certain relationships can reasonably be anticipated. Since both the CAC40 and the EuroStoxx50 represent European markets with significant overlap in large capitalization firms, we expect to find a strong and positive correlation between their returns. The evolution of the CAC40 is therefore likely to be closely linked to that of the EuroStoxx50.

In contrast, the Dow Jones and the S&P500 represent the United States market. Although they are geographically distant from the CAC40, their global influence suggests that movements in these indices may also affect the French market, but with a lower degree of correlation compared to the EuroStoxx50.

More specifically, the expected results are :

- A higher correlation between the CAC40 and the EuroStoxx50 returns.
- Moderate but significant correlations between the CAC40 and the U.S. indices (Dow Jones and S&P500).
- Regression models that confirm the EuroStoxx50 as the main explanatory factor for the CAC40 evolution, while U.S. indices may provide additional explanatory power.

These expectations will serve as a benchmark against which the empirical results will be compared.

1 Data Preparation

We begin by performing preliminary data preprocessing, reading the dataset using **pandas** to obtain a complete data frame. A first inspection reveals that the price series are non-stationary (Figure 1), making a direct regression on the raw prices a classical "textbook case" where linear regression assumptions are violated.

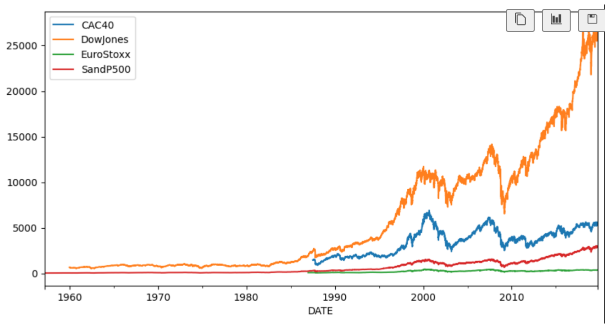


FIGURE 1 – Visualization of the stock prices of the indices analyzed.

Plotting the autocorrelation function (ACF) of order 1 would quickly reveal significant autocorrelation, indicating that a simple linear regression model on the price series would be poorly specified. In particular, regressing non-stationary series that are not cointegrated leads to :

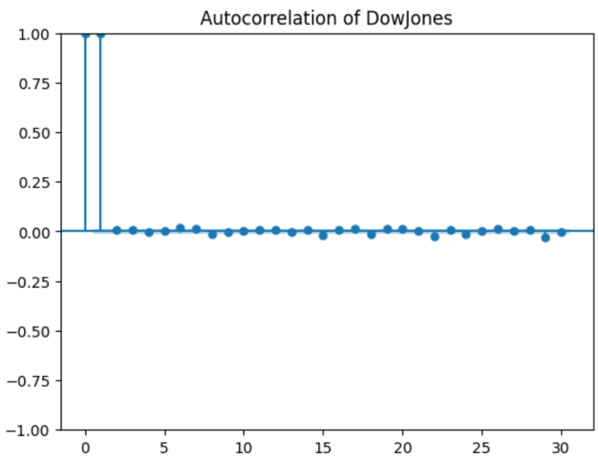
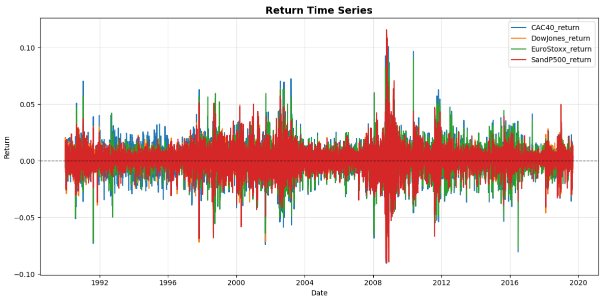


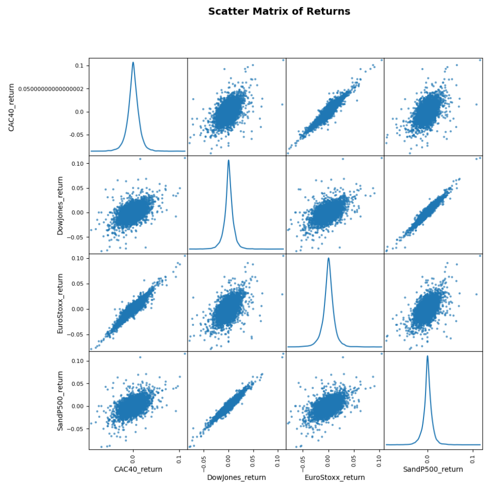
FIGURE 2 – Partial Autocorrelation of the Dow Jones

- **Spurious regression outcomes**, where coefficients appear significant and R^2 is high despite the absence of a true relationship.
- Residuals that exhibit strong **autocorrelation**, violating the classical assumption of independence and leading to biased standard errors and invalid hypothesis tests.
- General **model misspecification**, since the linear regression framework assumes stationary residuals.

To mitigate these issues, it is common practice to transform the price series into returns (or differences), which are generally stationary and better suited for regression analysis.



(a) Monthly Return of the different time series



(b) Pair Plot of each return Series

Commentary on Figures Figure 4a presents the yields (returns) of the different time series. At first glance, the distribution of the returns appears more suitable for linear modeling compared to the raw price series. The removal of trends and stabilization of variance through return computation makes it easier to imagine potential linear relationships between the indices.

Figure 4b shows a pair plot of all return series. The visualization suggests particularly strong linear relationships between the **CAC40** and the **EuroStoxx50**, as well as between the **S&P500** and the **Dow Jones**. These patterns indicate that changes in one index are closely mirrored by the other within each pair, supporting the use of linear regression models for further analysis.

TABLE 1 – Descriptive statistics of the return series from 1990 to 2023

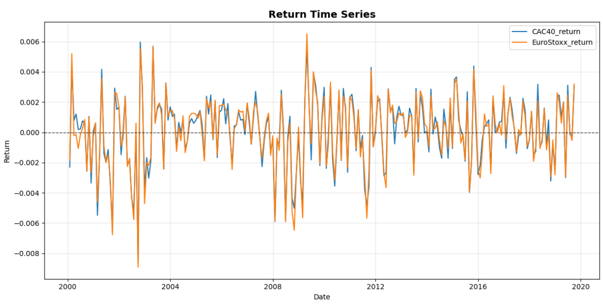
Series	Mean	Std	Min	Max	Cumulative Return
CAC40_yield	0.000222	0.013361	-0.090368	0.111762	1.805965
DowJones_yield	0.000348	0.010362	-0.078733	0.110803	8.799924
EuroStoxx_yield	0.000222	0.012004	-0.079184	0.104728	2.187280
SandP500_yield	0.000334	0.010826	-0.090350	0.115800	7.457131

Commentary on Descriptive Statistics The mean daily returns are small but positive for all indices, while volatility is slightly higher for the European indices (CAC40 and EuroStoxx50) compared to the U.S. indices (Dow Jones and S&P500). Extreme values highlight occasional market shocks. Cumulative returns show stronger growth for U.S. markets over the sample period. Overall, these returns are now suitable for linear regression analysis.

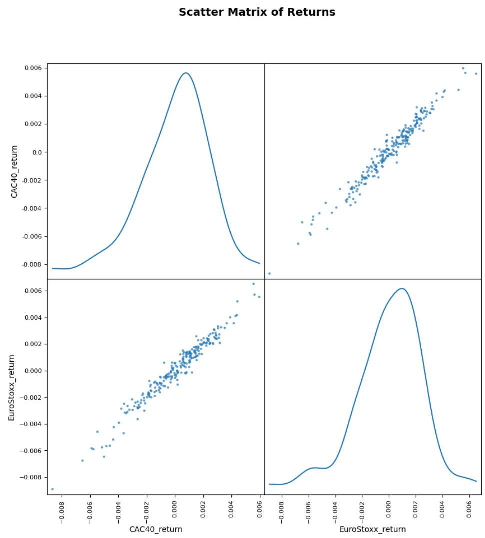
Transformation to Monthly Series and Preliminary Analysis

To reduce high-frequency noise and facilitate analysis over a longer horizon, the daily return series are converted into monthly series. Missing observations, if any, are removed to ensure a complete and consistent dataset. The resulting monthly database is referred to as **MIndexes**.

Preliminary visualization is then performed by plotting the monthly returns of two selected indices over a sub-period of interest, which allows for an initial inspection of trends and potential correlations.



(a) Monthly return of the different time series



(b) Pair plot of each return Series

2 Hypothesis on Data

Following the computation of returns, the data appear more suitable for linear regression analysis. The removal of trends and stabilization of variance through the calculation of returns ensures that the series are closer to satisfying the classical assumptions of the linear regression model.

Specifically, we will now conduct our analysis under the following standard assumptions of linear regression :

- **Linearity** : The relationship between the dependent variable (e.g., CAC40 returns) and the independent variables (other indices' returns) is assumed to be linear.
- **Independence of residuals** : The residuals of the regression should be uncorrelated across observations.
- **Homoscedasticity** : The variance of the residuals should be constant across all levels of the independent variables.
- **Normality of residuals** : The residuals are assumed to follow a normal distribution, which is especially relevant for hypothesis testing.

- **Stationarity of residuals** : Since the original series were non-stationary, it is crucial that the regression residuals are stationary to avoid spurious results.

By working with returns rather than raw prices, we aim to satisfy these assumptions, enabling meaningful estimation of the regression coefficients and valid statistical inference. The next step involves examining the relationships between indices through linear regression while systematically testing these hypotheses.

3 Linear Regression Analysis and Statistical Tests

Computation of Empirical Moments and Regression Parameters

To assess the linear relationships between the indices, a simple linear regression model is estimated using multiple approaches :

$$\text{CAC40\_yield} = \beta_0 + \beta_1 \cdot \text{EuroStoxx\_yield} + \varepsilon$$

1. **Manual computation using empirical sums :** Based on the computed sums, sums of squares, and cross-products, we calculate the empirical regression coefficients (β_0, β_1), the residual variance, and the coefficient of determination R^2. This approach provides an intuitive understanding of the relationships between variables.
2. **Matrix-based regression :** Using the **regression_matrix** and **linear_regression_with_matrix** functions, we estimate the regression coefficients for all pairs of yield series in matrix form. This allows for systematic computation and easy comparison across all indices.
3. **Standard OLS via statsmodels :** The **run_regression** function implements the ordinary least squares regression using the **statsmodels** library. This method provides the official regression output, including standard errors, t-statistics, p-values, and the full regression summary. The results are saved in **IndexRegResults.txt**.

Comparison across these three approaches shows highly consistent results :

- $\beta_0 \approx 0$, indicating an intercept not significantly different from zero.
- $\beta_1 \approx 1.03$–1.04, suggesting a near one-to-one linear relationship between EuroStoxx50 and CAC40 daily returns.
- $R^2 \approx 0.856$, meaning that approximately 85% of the variation in CAC40 returns is explained by EuroStoxx50 returns.
- Standard errors, t-statistics, and p-values confirm the slope is highly significant ($p < 0.001$), while the intercept is not significant.

This multi-method approach provides both a rigorous and transparent verification of the linear relationship between the indices, and ensures that the results are robust and reproducible.

Hypothesis Testing

After estimating a model, we see if we can simplify the computed models :

Intercept $(\beta_0 = 0)$ **:** The t-statistic is computed as :

$$t = \frac{\hat{\beta}_0}{\mathrm{SE}(\hat{\beta}_0)} \approx 0.149, \quad p \approx 0.882$$

Since $p \gg 0.05$, we fail to reject the null hypothesis. The intercept is not significantly different from zero.

Slope $(\beta_1 = 0)$ **:** The t-test for the slope confirms the high significance of the EuroStoxx50 coefficient in explaining CAC40 returns.

Analysis of Relationships Between Indices

Based on both the pairwise regressions and the scatter plots, we summarize the regression results in Figure 5, which displays the estimated coefficients (β_0, β_1), the coefficient of determination (R^2), and the results of hypothesis tests including p-values.

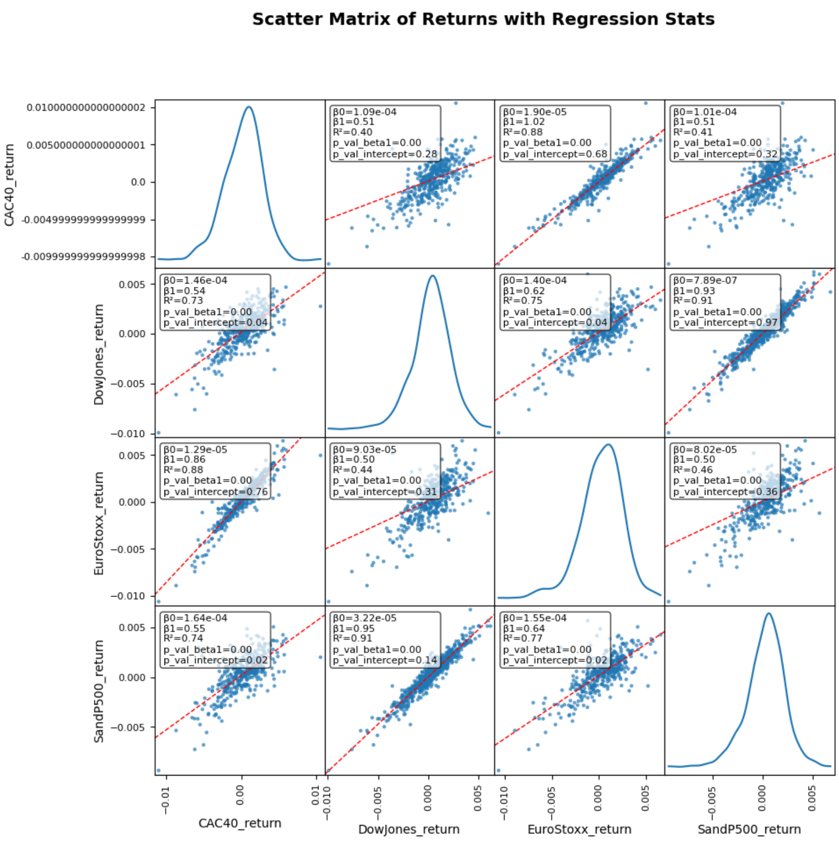


FIGURE 5 – Summary of OLS regression estimates for the return series.

- The CAC40 and EuroStoxx50 exhibit a very strong linear relationship, with $\beta_1 \approx 1$ and high R^2, indicating that European equity markets move closely together.
- The Dow Jones and S&P500 also show a strong linear dependency, reflecting the high co-movement within U.S. equity indices.
- Cross-market relationships (e.g., CAC40 vs. U.S. indices) are weaker but still positive, suggesting a global market effect where major indices tend to move in the same direction.

Conclusion

From our analysis of the regression relationships among the CAC40, EuroStoxx50, Dow Jones, and S&P500 yields, we can draw several key insights :

1. **Strong intra-European link :** The CAC40 is highly correlated with the EuroStoxx50, with a regression coefficient $\beta_1 \approx 1.04$ and $R^2 \approx 0.856$. This indicates that about 86% of the variance of the CAC40 can be explained by the EuroStoxx50, reflecting strong co-movement within European equity markets.
2. **Moderate influence from US indices :** The CAC40 also exhibits positive correlations with the Dow Jones and S&P500, with significant β_1 coefficients (approximately 0.40–0.41), but lower explanatory power ($R^2 \approx 0.21$˘0.22). While US markets do impact the CAC40, their effect is smaller compared to European indices.
3. **Very strong US market link :** The Dow Jones and S&P500 are extremely closely related, with $\beta_1 \approx 0.94$˘0.96 and $R^2 \approx 0.92$˘0.93, indicating that movements in one index almost entirely explain the variance of the other.
4. **Intercepts :** Across all regressions, the intercepts are not significantly different from zero, confirming that the relationships are mainly driven by the slopes (β_1).

Overall, these results highlight strong intra-regional co-movements, particularly within Europe and within the US, and demonstrate that global equity markets are interconnected, with European indices being most influential for the CAC40, and US indices being tightly linked among themselves.

Further Considerations

Although our analysis relied on standard linear regression, additional diagnostic tests could further validate the assumptions of the model :

- **Heteroskedasticity tests :** Tests such as Breusch-Pagan or White could be applied to check whether the variance of residuals is constant. Detecting heteroskedasticity is important, as it affects the reliability of standard errors, confidence intervals, and hypothesis tests.
- **Residual autocorrelation :** While returns are generally stationary, it would be informative to analyze residual autocorrelation using the Durbin-Watson statistic.

Significant autocorrelation would indicate that a simple linear model may be misspecified, suggesting potential benefits from using ARIMA or other time series models.

- **Robustness checks :** Alternative estimation generalized least squares could be employed to mitigate potential violations of homoscedasticity thereby strengthening inference on regression coefficients.

In conclusion, our multi-approach methodology—manual calculation, matrix-based regression, and OLS estimation—ensures both the validity and reproducibility of results, while providing a solid framework for further econometric investigations into the dynamics of equity markets.