**Assignment 1: Financial Time Series Analysis**

**Task 1: Multicollinearity in Financial Time Series Analysis**

**Introduction**

In econometrics, multicollinearity occurs when explanatory variables are highly correlated. This inflates standard errors, making it difficult to interpret individual regression coefficients, though the overall model may still fit well.  
  
The focus of this report is to investigate whether multicollinearity is a concern in financial time series regression. For this purpose, daily adjusted closing prices from Apple (AAPL), Microsoft (MSFT), and the S&P 500 Index (SPY) covering January 2019 to January 2024 were obtained from Yahoo Finance.  
  
The data were transformed into logarithmic returns to ensure stationarity and comparability. A regression model was estimated where AAPL returns were explained by MSFT and SPY returns, with diagnostics including correlation matrices, OLS regression output, and Variance Inflation Factors (VIF) used to assess multicollinearity.

**Empirical Findings**

1. **Correlation Heatmap :** The correlation heatmap shows strong pairwise correlations:

* AAPL–MSFT: ρ = 0.76
* AAPL–SPY: ρ = 0.81
* MSFT–SPY: ρ = 0.83

This indicates strong co-movement and signals multicollinearity.

1. **OLS Regression Results :** The regression of AAPL returns on MSFT and SPY yields:

* R² = 0.678 (model explains 67.8% of AAPL variation)
* MSFT coefficient = 0.3172 (p < 0.001)
* SPY coefficient = 0.8536 (p < 0.001)
* Condition Number = 156 (indicates collinearity)

1. **Variance Inflation Factor (VIF)**

Interpretation: MSFT and SPY both have VIF ≈ 3.1, suggesting moderate but not severe multicollinearity.

1. **Line Graphs :** Two graphs were used to evaluate predictive accuracy:

* Daily Returns: Predicted vs. actual AAPL returns track closely, though volatility spikes create deviations.
* Cumulative Returns: Predicted series follows the overall trend, but underestimates AAPL’s firm-specific rallies.

**Discussion**

Multicollinearity makes inference difficult: it is hard to determine whether MSFT or SPY is the primary driver of AAPL’s returns. However, prediction remains strong since both variables together capture market-wide movements.  
  
The high condition number, strong correlations, and moderate VIF confirm multicollinearity is present. Although not severe enough to invalidate the model, it affects interpretation. Common econometric remedies include Principal Component Analysis (PCA), factor models, or ridge regression, which reduce dimensionality and mitigate collinearity.

**Conclusion**

This analysis shows that multicollinearity is a genuine concern in financial time series regressions. While it does not bias estimates, it inflates variances and undermines inference. These results confirm that multicollinearity is present in the regression of AAPL on MSFT and SPY. Although not severe enough to invalidate the model, it complicates inference. For forecasting purposes, this degree of multicollinearity is unlikely to pose a major problem.

**Task 2: Time Series Regression on S&P 500 Index**

**Model Specification and Estimation**

We regressed the annual level of the S&P 500 Index (1980–2023) on Year to examine its long-term behavior:  
  
 SP500\_t = β₀ + β₁Year\_t + ε.

**Regression Results Interpretation**

The regression yielded a significant positive slope (β ≈ 78.5, p < 0.001) and R² = 0.798, suggesting a strong fit.

However, Durbin-Watson = 0.168 indicated strong autocorrelation, and the condition number (315,000) pointed to trend-induced collinearity. The fitted regression line under predicted recent exponential growth and implied negative index levels in the early 1980s. This is consistent with a spurious regression.

**Identification of Issues**

The model is economically implausible because it assumes linear growth with time. Both Year and SP500 are trending non-stationary variables, producing artificial correlation. OLS assumptions are violated due to serial correlation and non-stationarity.

**Proposed Resolution**

To address these issues:

* Transform index levels into annual log returns (stationary series).
* Estimate models using ARIMA or GARCH to account for autocorrelation and volatility clustering.
* For long-run analysis, use co-integration techniques to link the S&P 500 with economic fundamentals such as GDP or interest rates.

**Conclusion**

The regression of SP500 levels on Year appears statistically strong but is spurious. The high R² arises from shared upward trends, not a causal relationship. Transforming data into returns provides a more realistic modeling framework, ensuring stationarity and avoiding misleading inference.

**References**

1. Brooks, C. (2019). *Introductory econometrics for finance* (4th ed.). Cambridge University Press.
2. Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (5th ed.). McGraw-Hill.
3. Hamilton, J. D. (1994). *Time series analysis*. Princeton University Press.
4. Tsay, R. S. (2010). *Analysis of financial time series* (3rd ed.). Wiley.
5. Wooldridge, J. M. (2016). *Introductory econometrics: A modern approach* (6th ed.). Cengage Learning.

**Appendix: Empirical Analysis Output**

**Task 1 Code and results**

**Task 2 Code and Results**