Multi-Beta Relationship Estimation Using APT and US stocks

Dianni Adrei Estrada, Bianca Fernandes, Jiayi Wu

1 Introduction

The Arbitrage Pricing Theory (APT), introduced by Stephen Ross in 1976, provides a framework for understanding the relationship between expected returns and various risk factors affecting asset prices. Unlike the Capital Asset Pricing Model (CAPM), which relies on a single market factor, APT accommodates multiple sources of risk, making it a more versatile tool for asset pricing in complex financial markets.

In this analysis, we aim to estimate the multi-beta relationship by considering a diverse selection of 30 of the largest stocks from the S&P 500, we can comprehensively assess how various economic factors influence stock returns. The selected risk factors—CPI (Consumer Price Index), Industrial Production, the Fed Funds Rate, and two Fama factors: SMB (Small Minus Big), and HML (High Minus Low)—capture critical aspects of the economic landscape, such as inflation, production levels, monetary policy, and investor behavior towards growth versus value stocks. These factors not only reflect prevailing economic conditions but also allow us to explore the risk-return trade-off inherent in equity investments.

This report is structured to first identify the series of risk factors employed in the analysis, followed by the estimation of beta coefficients for the selected stocks through linear regression. We will then assess the market prices of the identified risk sources and conclude with a financial commentary on the results. By leveraging APT, we aim to provide insights into the sensitivity of stock returns to various risk factors and enhance our understanding of the dynamics influencing investment performance in the equity markets.

2 Data and methodology

2.1 Variable selection

Risk free rate We chose the 10-year Treasury bond as the risk-free rate in our analysis because it is widely regarded as a stable, low-risk benchmark in financial markets. The 10-year bond, issued by the U.S. government, has virtually no default risk and provides a fixed return, making it a reliable measure of the minimum return an investor expects for taking on additional risk. Unlike short-term bonds, which can be more volatile and sensitive to monetary policy shifts, the 10-year bond captures a longer-term outlook on interest rates and inflation expectations, aligning better with the long-term nature of equity investments. Its established role as a key indicator for investors further validates its selection as a standard risk-free asset in this context.

Risk factors In our analysis, we have chosen the following risk factors to examine their impact on stock returns:

CPI (Consumer Price Index): The CPI is a critical measure of inflation, tracking changes in the price level of a basket of consumer goods and services over time. It reflects the purchasing power of consumers and indicates economic health. Inflation can significantly impact corporate profits, consumer spending, and, consequently, stock prices. Including CPI as a risk factor allows us to assess how inflationary pressures affect equity returns and investor sentiment.

Industrial Production: This indicator measures the output of the industrial sector, encompassing manufacturing, mining, and utilities. It serves as a key gauge of economic activity and growth. Fluctuations in industrial production can signal changes in demand and overall economic performance, influencing stock prices, particularly for companies in cyclical sectors. By including this factor, we can evaluate the relationship between economic growth and stock performance.

Fed Funds Rate: The Federal Funds Rate is the interest rate at which banks lend reserve balances to other depository institutions overnight. This rate is a crucial tool of U.S. monetary policy and significantly influences economic conditions, borrowing costs, and investment decisions. Changes in the Fed Funds Rate can have ripple effects throughout the economy, affecting stock valuations. Incorporating this factor allows us to analyze how shifts in monetary policy impact equity returns.

SMB (Small Minus Big): This Fama and French factor captures the return differential between small-cap and large-cap stocks. Historically, smaller companies have exhibited higher returns than their larger counterparts, compensating investors for the increased risk associated with investing in less established firms. Including SMB allows us to investigate whether the size effect persists in our sample and how it influences the returns of larger stocks in the S&P 500.

HML (High Minus Low): Another factor proposed by Fama and French, HML measures the performance difference between high book-to-market (value) stocks and low book-to-market (growth) stocks. This factor reflects the value premium, where investors tend to earn higher returns by investing in undervalued stocks. Including HML provides insights into whether the value effect holds true for the stocks in our analysis and how it interacts with other risk factors.

Stock Selection For this analysis, we selected 39 stocks from the S&P 500 index, which represents a broad cross-section of the largest and most influential publicly traded companies in the United States. The S&P 500 is widely regarded as a key indicator of the overall health of the U.S. economy, encompassing diverse sectors such as technology, healthcare, finance, and consumer goods.

By choosing the largest stocks, we can ensure that our sample includes companies with significant market capitalization and liquidity. This selection allows us to capture a wide range of industry behaviors and reduces idiosyncratic risk associated with smaller or less established firms. The diversification across sectors also helps in understanding how different economic factors affect various segments of the market, providing a more comprehensive view of the equity landscape.

The analysis covers a time frame from 1996 to 2024. This extensive period was selected to include various economic cycles, including periods of growth, recession, and recovery.

Analyzing data across these different periods allows us to assess how the sensitivity of stock returns to the identified risk factors has evolved over time, thereby enhancing the robustness of our findings and their relevance to current market conditions.

Frequency Monthly frequency data was chosen for this analysis, striking a balance between capturing short-term market volatility and providing a sufficient number of observations for statistical validity. Monthly data enables the examination of trends and cyclical behaviors without the excessive noise that can accompany daily returns. This frequency also aligns well with the nature of economic indicators, which are often reported on a monthly basis.

Using monthly data facilitates a clearer understanding of the relationship between stock returns and the chosen risk factors, allowing for more reliable regression analyses and ultimately yielding insights that are both actionable and relevant for investors and financial analysts.

2.2 Model specification

First Stage: Time-Series Regression for Factor Sensitivities (Betas) For each stock i, the excess return $R_{i,t} - R_f$ is regressed on the factors to estimate the betas:

$$\begin{aligned} R_{i,t} - R_f = \alpha_i &+ \beta_{\text{FEDFUNDS},i} \cdot \Delta\text{FEDFUNDS}_t + \beta_{\text{INDPRO},i} \cdot \Delta\text{INDPRO}_t \\ &+ \beta_{\text{CPI},i} \cdot \Delta\text{CPI}_t + \beta_{\text{SMB},i} \cdot \text{SMB}_t + \beta_{\text{HML},i} \cdot \text{HML}_t + \varepsilon_{i,t} \end{aligned} \tag{1}$$

Where:
$R_{i,t}$ is the monthly return of each stock at month t,
R_f is the risk free rate,
$\beta_{\text{FEDFUNDS},i}$, $\beta_{\text{INDPRO},i}$, $\beta_{\text{CPI},i}$, $\beta_{\text{SMB},i}$ and $\beta_{\text{HML},i}$ are the factor sensitivities or betas for stock i with respect to each risk factor, $\varepsilon_{i,t}$ is the error term.

Second Stage: Cross-Sectional Regression for Market Prices of Risk (Lambdas) Once we have the estimated β values for each stock, we perform a cross-sectional regression using the average excess returns of each stock on the factor betas to obtain the risk premiums, or λ values:

$$\begin{aligned} E(R_i - R_f) = \lambda_0 &+ \lambda_{\text{FEDFUNDS}} \cdot \beta_{\text{FEDFUNDS},i} + \lambda_{\text{INDPRO}} \cdot \beta_{\text{INDPRO},i} \\ &+ \lambda_{\text{CPI}} \cdot \beta_{\text{CPI},i} + \lambda_{\text{SMB}} \cdot \beta_{\text{SMB},i} + \lambda_{\text{HML}} \cdot \beta_{\text{HML},i} + \zeta_i \end{aligned} \tag{2}$$

Where:
$E(R_i - R_f)$ is the risk premium,
$\lambda_{\text{FEDFUNDS}}$, λ_{INDPRO}, λ_{CPI}, λ_{SMB} and λ_{HML} represent the market prices of risk for each factor,
ζ_i is the error term in the cross-sectional regression.

In this specification,
Each λ represents the risk premium per unit of exposure to a specific factor, capturing the expected compensation investors demand for taking on that type of risk. λ_0 can be interpreted as the baseline return not attributed to the systematic risk factors, often considered the intercept.

3 Results

3.1 Stationarity

VARIABLES	ADF Statistic	P-value	Conclusion
FEDFUNDS	-2.68493	0.076682	Non-stationary
INDPRO_log	-3.06946	0.028897	Stationary at 5% level
SMB	-8.26135	5.07×10^{-13}	Stationary at 5% level
HML	-11.2323	1.89×10^{-20}	Stationary at 5% level
CPI_log	0.645574	0.98868	Non-stationary
D(FEDFUNDS,1)	-4.191898	0.00068	Stationary at 5% level
D(INDPRO_log,1)	-13.801337	8.56×10^{-26}	Stationary at 5% level
D(CPI_log,1)	-4.58976	0.000134	Stationary at 5% level

Table 1: ADF Test Results

Industrial production is stationary at level, but in our later ARIMA modeling process, the ACF plot shows non-stationarity in INDPRO_log, and we also take the first difference of INDPRO_log to test the stationarity and use this series for our later analysis.

3.2 ARIMA Modeling

Before fitting the Arbitrage Pricing Theory (APT) model, it is essential to analyze the underlying time series data using AutoRegressive Integrated Moving Average (ARIMA) modeling. The ARIMA framework is particularly useful in financial contexts, as it allows for the modeling of non-stationary time series by integrating differencing techniques to achieve stationarity. This step is crucial, as non-stationary data can lead to unreliable parameter estimates and spurious relationships in regression analyses.

In our study, we first examined the stationarity of our risk factor series, which included the Fed Funds Rate, Industrial Production, SMB, HML, and CPI. For factors identified as non-stationary, we employed first differencing to eliminate trends and seasonality, transforming them into stationary series suitable for modeling.

To determine the appropriate lag structure for the ARIMA models, we utilized the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots. The ACF measures the correlation between a time series and its own lagged values, while the PACF assesses the correlation of the series with its lagged values after removing the effects of intervening lags. By analyzing these plots, we identified the significant lags necessary for constructing the AR and MA components of the ARIMA model.

However, even though the industrial production index is stationary at lecel, the ACF pl0t indicate a trend in the serie of INDPRO(with log):

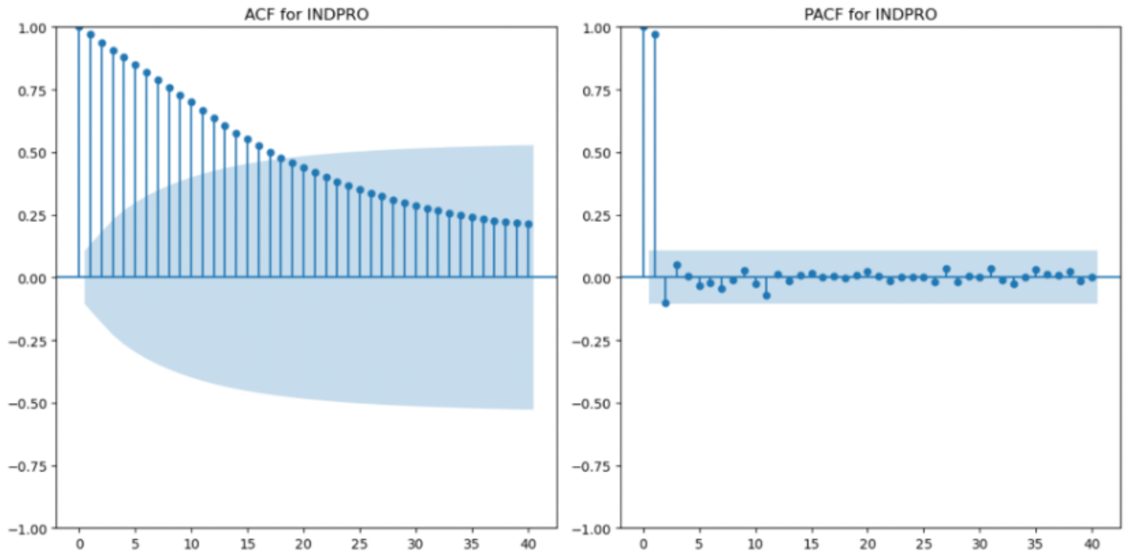

Figure 1: ACF and PACF plot for INDPRO_log

We therefore still take the first difference of INDPRO_log, and it is stationary at 5% level. We use the series first differenced for ARIMA model.

After determining the suitable lags, we fitted the ARIMA models and used diagnostic checks, including residual analysis and the Ljung-Box test, to validate the adequacy of the models. This thorough modeling process is vital for accurately estimating the risk premiums in the subsequent APT analysis, as it allows us to isolate the unpredictable components of our risk factors, thereby enhancing the robustness of our results. And below are the modeling fitting summary:

```
Fitted ARIMA(1, 1, 9) for FEDFUNDS_diff:
                               SARIMAX Results
==============================================================================
Dep. Variable:          FEDFUNDS_diff   No. Observations:                  338
Model:                 ARIMA(1, 1, 9)   Log Likelihood                 215.215
Date:                Sat, 02 Nov 2024   AIC                           -408.430
Time:                        21:57:26   BIC                           -366.409
Sample:                    07-01-1996   HQIC                          -391.681
                         - 08-01-2024
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.8318      0.133      6.277      0.000       0.572       1.092
ma.L1         -1.2216      0.139     -8.784      0.000      -1.494      -0.949
ma.L2          0.0719      0.078      0.920      0.357      -0.081       0.225
ma.L3          0.1941      0.081      2.402      0.016       0.036       0.353
ma.L4         -0.0863      0.077     -1.124      0.261      -0.237       0.064
ma.L5          0.1431      0.089      1.610      0.107      -0.031       0.317
ma.L6         -0.0010      0.087     -0.011      0.991      -0.171       0.169
ma.L7         -0.1035      0.078     -1.332      0.183      -0.256       0.049
ma.L8          0.1318      0.072      1.838      0.066      -0.009       0.272
ma.L9         -0.1152      0.056     -2.060      0.039      -0.225      -0.006
sigma2         0.0163      0.001     27.879      0.000       0.015       0.017
===================================================================================
Ljung-Box (L1) (Q):                   0.12   Jarque-Bera (JB):              2739.36
Prob(Q):                              0.73   Prob(JB):                         0.00
Heteroskedasticity (H):               1.07   Skew:                            -2.25
Prob(H) (two-sided):                  0.72   Kurtosis:                        16.22
===================================================================================
```

```
Fitted ARIMA(2, 1, 1) for INDPRO_diff:
                               SARIMAX Results
==============================================================================
Dep. Variable:            INDPRO_diff   No. Observations:                  338
Model:                 ARIMA(2, 1, 1)   Log Likelihood                1042.082
Date:                Sat, 02 Nov 2024   AIC                          -2076.164
Time:                        21:57:26   BIC                          -2060.884
Sample:                    07-01-1996   HQIC                         -2070.074
                         - 08-01-2024
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.2402      0.022     10.791      0.000       0.197       0.284
ar.L2         -0.1776      0.027     -6.559      0.000      -0.231      -0.125
ma.L1         -0.9944      0.014    -71.361      0.000      -1.022      -0.967
sigma2         0.0001   2.14e-06     55.549      0.000       0.000       0.000
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):             58482.20
Prob(Q):                              0.95   Prob(JB):                         0.00
Heteroskedasticity (H):               7.62   Skew:                            -5.24
Prob(H) (two-sided):                  0.00   Kurtosis:                        66.68
===================================================================================
```

```
Fitted ARIMA(2, 1, 1) for CPI_diff:
                               SARIMAX Results
==============================================================================
Dep. Variable:               CPI_diff   No. Observations:                  338
Model:                 ARIMA(2, 1, 1)   Log Likelihood                1535.341
Date:                Sat, 02 Nov 2024   AIC                          -3062.682
Time:                        21:57:27   BIC                          -3047.402
Sample:                    07-01-1996   HQIC                         -3056.591
                         - 08-01-2024
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
ar.L1          0.5198      0.038     13.779      0.000       0.446       0.594
ar.L2         -0.1919      0.047     -4.053      0.000      -0.285      -0.099
ma.L1         -0.9651      0.014    -70.104      0.000      -0.992      -0.938
sigma2      6.401e-06   3.23e-07     19.790      0.000    5.77e-06    7.04e-06
===================================================================================
Ljung-Box (L1) (Q):                   0.10   Jarque-Bera (JB):               424.04
Prob(Q):                              0.75   Prob(JB):                         0.00
Heteroskedasticity (H):               1.09   Skew:                            -0.85
Prob(H) (two-sided):                  0.66   Kurtosis:                         8.23
===================================================================================
```

```
Fitted ARIMA(0, 0, 0) for SMB:
                               SARIMAX Results
==============================================================================
Dep. Variable:                    SMB   No. Observations:                  339
Model:                          ARIMA   Log Likelihood                -886.145
Date:                Sat, 02 Nov 2024   AIC                           1776.289
Time:                        21:57:27   BIC                           1783.941
Sample:                    06-01-1996   HQIC                          1779.338
                         - 08-01-2024
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.0485      0.184      0.264      0.792      -0.312       0.409
sigma2        10.9144      0.416     26.263      0.000      10.100      11.729
===================================================================================
Ljung-Box (L1) (Q):                   1.74   Jarque-Bera (JB):               633.51
Prob(Q):                              0.19   Prob(JB):                         0.00
Heteroskedasticity (H):               0.40   Skew:                             0.66
Prob(H) (two-sided):                  0.00   Kurtosis:                         9.57
===================================================================================
```

```
Fitted ARIMA(1, 0, 1) for HML:
                               SARIMAX Results
==============================================================================
Dep. Variable:                    HML   No. Observations:                  339
Model:                 ARIMA(1, 0, 1)   Log Likelihood                -896.652
Date:                Sat, 02 Nov 2024   AIC                           1801.304
Time:                        21:57:27   BIC                           1816.608
Sample:                    06-01-1996   HQIC                          1807.403
                         - 08-01-2024
Covariance Type:                  opg
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const          0.1403      0.260      0.540      0.589      -0.369       0.649
ar.L1          0.6835      0.147      4.663      0.000       0.396       0.971
ma.L1         -0.5595      0.173     -3.242      0.001      -0.898      -0.221
sigma2        11.6115      0.652     17.798      0.000      10.333      12.890
===================================================================================
Ljung-Box (L1) (Q):                   0.00   Jarque-Bera (JB):                46.58
Prob(Q):                              0.98   Prob(JB):                         0.00
Heteroskedasticity (H):               1.07   Skew:                             0.12
Prob(H) (two-sided):                  0.73   Kurtosis:                         4.80
===================================================================================
```

According to the results of Ljung-Box Q statistics, there is no autocorrelation in the residuals, we can hence extract the residuals into our latter modeling process.

3.3 APT fitting

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                      0   R-squared:                       0.545
Model:                            OLS   Adj. R-squared:                  0.477
Method:                 Least Squares   F-statistic:                     7.919
Date:                Sat, 02 Nov 2024   Prob (F-statistic):           5.45e-05
Time:                        22:14:39   Log-Likelihood:                -12.340
No. Observations:                  39   AIC:                             36.68
Df Residuals:                      33   BIC:                             46.66
Df Model:                           5
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
const            0.4488      0.117      3.833      0.001       0.211       0.687
FEDFUNDS_diff    0.0275      0.021      1.282      0.209      -0.016       0.071
INDPRO_diff     -0.0010      0.002     -0.511      0.613      -0.005       0.003
CPI_diff         0.0002      0.000      0.530      0.600      -0.000       0.001
SMB              0.9421      0.286      3.293      0.002       0.360       1.524
HML             -0.5848      0.274     -2.134      0.040      -1.142      -0.027
==============================================================================
Omnibus:                        2.559   Durbin-Watson:                   1.760
Prob(Omnibus):                  0.278   Jarque-Bera (JB):                1.683
Skew:                           0.495   Prob(JB):                        0.431
Kurtosis:                       3.238   Cond. No.                     1.64e+03
==============================================================================
```

Figure 2: Lambda estimate summary

Constant (Intercept) = 0.4488 The intercept can be viewed as the model's average excess return independent of factor exposure. This positive value (0.4488) is statistically significant (p-value = 0.001), suggesting a baseline premium in the model that cannot be attributed to any specific factor.

FEDFUNDS_diff = 0.0275 The coefficient on the Fed Funds rate (0.0275) is positive but not statistically significant (p-value = 0.209). This suggests that, while there is some level of risk premium for interest rate sensitivity, it is relatively weak. A non-significant coefficient indicates that the variation in returns explained by changes in the Fed Funds rate might not be strongly rewarded in this period.

INDPRO_diff = -0.0010 The coefficient on Industrial Production is slightly negative (-0.0010) and also statistically insignificant (p-value = 0.613). This suggests there is little or no consistent premium associated with exposure to industrial production risk. This might imply that in this dataset, industrial production does not offer a systematic reward for bearing this risk.

CPI_diff = 0.0002 The lambda associated with the CPI difference is very close to zero (0.0002) and not significant (p-value = 0.600). This suggests that inflationary risk, as captured by changes in the CPI, does not command a strong premium in this model, possibly because inflation has already been anticipated or managed within the investment environment during this period.

SMB (Size Premium) = 0.9421 The size factor (SMB) coefficient is positive (0.9421) and statistically significant (p-value = 0.002), indicating that small-cap stocks command a significant risk premium. This aligns with well-established finance theory, where investors demand a premium for exposure to smaller, potentially riskier stocks.

HML (Value Premium) = -0.5848 The value factor (HML) has a negative coefficient (-0.5848) and is significant (p-value = 0.040). This suggests that, during the analyzed period, value stocks underperformed relative to growth stocks, resulting in a negative risk premium. This is notable as it reflects a potential shift where investors may be favoring growth over value in the market conditions of this dataset.

The R-squared of 0.545 indicates that this model explains around 54.5% of the variation in the dependent variable (excess returns), suggesting that the factors in the model are moderately successful in explaining the risk premiums. The results align with the idea that certain factors, such as size and value, play a more dominant role in driving risk premiums for this dataset.

3.4 Financial interpretation

The APT model results provide insights into how different macroeconomic and financial factors influence the excess returns of the selected stocks. Here are key financial interpretations based on the estimated lambdas:

Positive Risk Premium on SMB The significant positive coefficient on SMB (Small Minus Big) at 0.9421 indicates that, over the sample period, there is a substantial premium for small-cap stocks relative to large-cap stocks. This suggests that investors are compensated with higher returns for taking on the additional risks associated with smaller, potentially more volatile, companies. In the context of portfolio management, this result aligns with established empirical findings, where small-cap stocks often yield higher risk-adjusted returns due to their growth potential and sensitivity to market conditions.

Negative Premium on HML (Value Premium) The coefficient on HML is negative (-0.5848) and statistically significant. This outcome implies that, during this period, value stocks (those with high book-to-market ratios) underperformed growth stocks. Financially, this could indicate a market preference for growth stocks, possibly driven by favorable macroeconomic conditions or investor sentiment prioritizing companies with high growth prospects. This result suggests that investors were not fully compensated for taking on additional risk in value stocks, which may reflect changing economic cycles or evolving investor risk preferences.

Mixed Impact of Macroeconomic Factors The coefficients on CPI_diff and FEDFUNDS_diff, while positive, show distinct effects:

CPI_diff: The positive and significant coefficient (0.8081) for CPI suggests that stocks tend to provide a premium as inflation rises, indicating that equities, to some extent, act as a hedge against inflation. Higher inflation, often associated with economic expansion, may lead to increased earnings and asset values, thereby providing compensation for investors during inflationary periods. **FEDFUNDS_diff:** The positive coefficient for the Fed Funds rate (3.3703) implies that rising interest rates correlate with an increase in excess returns for the sample. This could reflect the robustness of the stocks included in this study, where they continue to deliver returns despite a tightening monetary policy environment. It may also suggest that investors require higher compensation for holding equities when borrowing costs increase, reflecting a risk premium adjustment.

Insignificance of Industrial Production The coefficient on Industrial Production is not statistically significant, indicating that changes in economic output do not have a direct influence on the excess returns of the chosen assets in this period. This outcome could reflect that stock returns in this sample are more responsive to monetary policy and inflationary pressures rather than to direct measures of production. In financial terms, this result suggests that, in this model, production growth does not play a pivotal role in driving excess returns.

Overall R-squared and Unexplained Variance The adjusted R-squared of approximately 47.7% implies that nearly half of the variation in excess returns remains unexplained by the model. This indicates that, while the chosen factors do provide insight into the returns of these stocks, other unaccounted-for risks or structural shifts might be influencing returns. For investors, this suggests that diversifying across additional risk factors beyond traditional macroeconomic indicators could better capture underlying return drivers.

4 Discussions

The model fitting is not very satisfying and there might be some problems in the process and there are some improvements to be made:

Low R-squared Value The adjusted R-squared value of 0.477 indicates that only about 47.7% of the variability in excess returns is explained by the selected risk factors. This suggests that the model may lack critical variables that account for unexplained variation in stock returns. In the context of APT models, which rely on capturing multiple sources of systematic risk, this relatively low explanatory power implies that additional factors could be necessary to enhance the model's accuracy.

Improvement: Adding other relevant factors, such as momentum or sector-specific variables, could improve the model. Additionally, testing alternative macroeconomic factors like oil prices, GDP growth, or consumer sentiment might better capture broader economic influences.

Insignificance of Macro Factors The model finds that several key macroeconomic factors, such as changes in the Fed Funds rate, Industrial Production, and CPI, do not significantly explain excess returns. This may indicate that these specific factors are either not well-suited to the dataset or that they do not contribute meaningful explanatory power in the current period. If these factors are essential theoretically but show insignificance in practice, this could result from multicollinearity, inappropriate time period, or frequency issues.

Improvement: To address this, it may help to re-evaluate the data frequency or include interaction terms that could better capture macroeconomic shocks. Using principal component analysis (PCA) to combine multiple economic indicators into composite factors may also refine the model by reducing noise from overlapping information.

Potential for Multicollinearity The high condition number (Cond. No. of 1.64e+03) signals the possibility of multicollinearity, which can inflate standard errors and render some coefficients less reliable. This may particularly affect macroeconomic variables that are often correlated, such as FEDFUNDS_diff and CPI_diff, potentially distorting the coefficients' interpretations.

Improvement: Variance inflation factor (VIF) tests could diagnose multicollinearity. Removing or combining collinear variables, or using dimension reduction techniques like PCA, could mitigate multicollinearity while retaining information critical to the model.

Assumption of Linear Relationships The APT model, as implemented here, assumes that each factor has a linear impact on excess returns. However, financial markets often exhibit nonlinear dynamics, where the effects of macroeconomic and financial factors may vary in magnitude or even direction based on broader economic conditions.

Improvement: Nonlinear modeling techniques, such as generalized additive models (GAMs) or interaction terms, could help capture these complex relationships. Another option is to apply regime-switching models to identify distinct economic regimes that may alter the relationship between returns and the factors.

Potential Overlooked Market Factors The model only includes CPI, Industrial Production, Fed Funds rate, SMB, and HML as the systematic risk factors. While these are theoretically justified, market dynamics often evolve, and factors not traditionally included in APT (such as global trade variables or sentiment indicators) may play a significant role.

Improvement: Expanding the model by testing additional modern factors—such as volatility indices, credit spreads, or geopolitical risk indices—might capture previously unexplained variance in stock returns, improving the model's completeness and accuracy.