UNIVERSITÉ PARIS 1
PANTHÉON SORBONNE

EMPIRICAL APPLICATION 3
Asset Pricing

Charaf ZGUIOUAR
Majd RABBAJ
Meriem BENDIR

Choice of our Variables, Period Selection and Frequency Selection

For our empirical application, we decided to choose data of: S&P500, US Real GDP and US 10-Yiear Treasury yield.

The choice of the S&P 500 as the stock index for this analysis is justified by its comprehensive market representation, its role as an economic indicator, its global significance, the availability of extensive historical data, and its utility for benchmarking and comparability purposes.

Period Selection (1985 to 2023): The period from 1985 to 2023 is selected for encompassing major financial events, including the 1987 crash, the dot-com bubble, the 2008 crisis, and the COVID-19 pandemic impacts. This extended timeframe allows for an understanding of market and economic fundamentals' responses to various economic cycles and crises, providing a comprehensive view of both short-term and long-term market trends.

Frequency Selection (Quarterly Data): Quarterly data is chosen to balance detail and overview, capturing significant market changes while avoiding the noise of more granular data. Aligning with the quarterly release of key economic reports like GDP ensures a consistent analysis that reflects broader economic trends

Checking that our series are I(1)

We aim at estimating the long-term fundamental value of the S&P500 index using the Present Value Model, expressed in its simplest form in the Gordon-Shapiro formula.

This identification holds significant value for investment strategies for two primary reasons:

Benchmark for Current Valuation: The fundamental value acts as a crucial reference point for evaluating the current market valuation of the index. This serves as a vital tool for implementing contrarian investment strategies, which involve making investment decisions that go against prevailing market trends, and for engaging in value investing, where the focus is on selecting stocks that are currently undervalued but are expected to rise in value in the future.

Differentiating Between Transitory and Long-Term Price Changes: Understanding the fundamental value of the index is key to distinguishing between short-term price fluctuations, often caused by temporary shocks, and genuine long-term shifts in the fundamental value of the index. This knowledge is instrumental in making informed investment decisions that account for both short-term market movements and long-term value trends.

The Gordon-Shapiro model suggest two fundamentals: dividends and a discount rate factor, specified as a risk-free rate plus an ex-ante risk premium as show in the Equation below:

$$\boldsymbol{P_t = \frac{D_t}{R^e{}_t - g^e{}_t}}$$

With :

P_t : Price of the stock at time t;
D_t : Dividends at distributed by the stock at time t;
$R^e{}_t$: The expected stock return;
$g^e{}_t$: The expected dividend growth rate

We consider the expected return $R^e{}_t$ as the sum of a risk-free return $r^F{}_t$ and ex-ante risk premium $\pi^e{}_t$:

$$\boldsymbol{P_t = \frac{D_t}{r^F{}_t + \Pi^e{}_t - g^e{}_t}}$$

Then we take the logarithm:

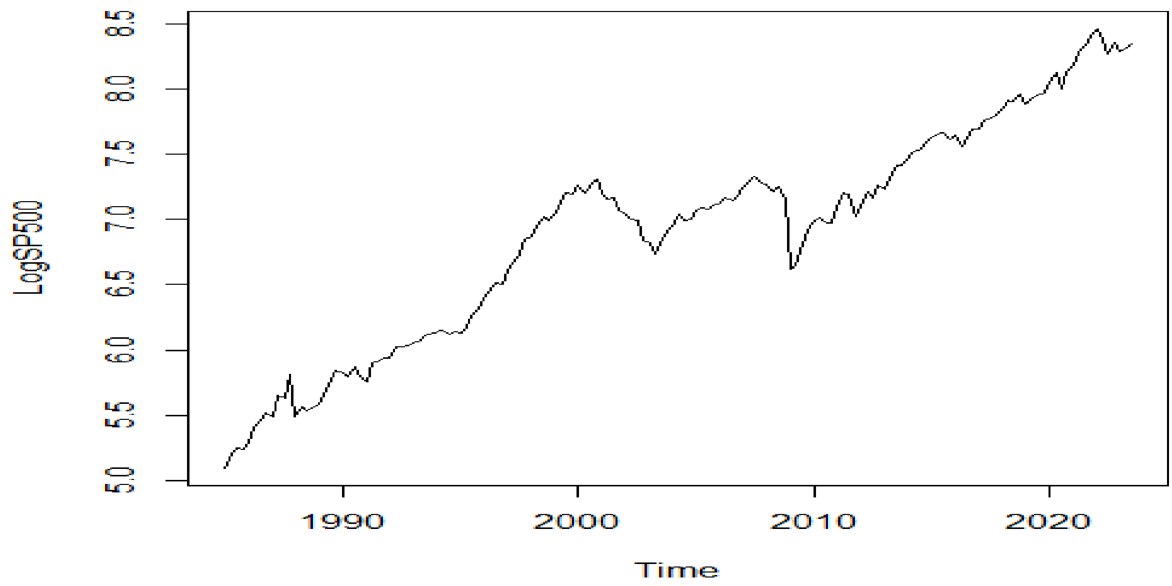
$$\boldsymbol{log(P_t) = log(D_t) - log(r^F{}_t + \Pi^e{}_t - g^e{}_t)}$$

Now, for the long-term equation describing the equilibrium price of the S&P500, we use:
The logarithm of the real GDP of the USA as a proxy for dividends $\log(D_t)$ as It provides a good approximation of the activity of the firms
The 10 year US Treasury bond yields as a proxy for the risk-free rate $r^F{}_t$

First, to begin our analysis we created our DataFrame with the values of my 3 series between 1985 and 2023 with Quarterly Data.

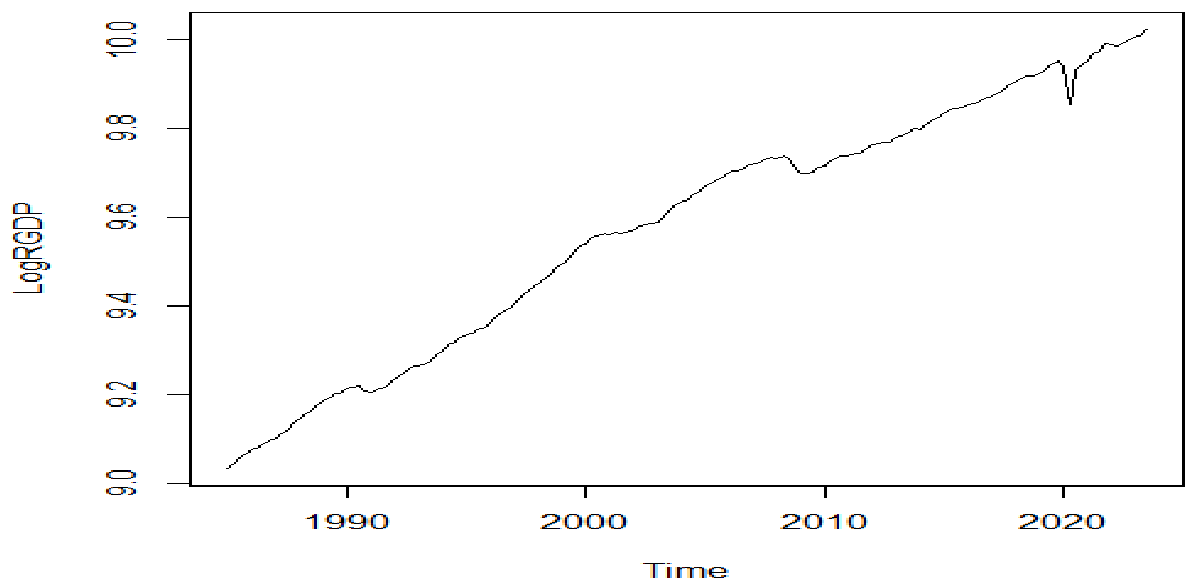
We plotted them as below and we can clearly see that the series does not appear to be stationary:

Chart 1: Logarithm of S&P500



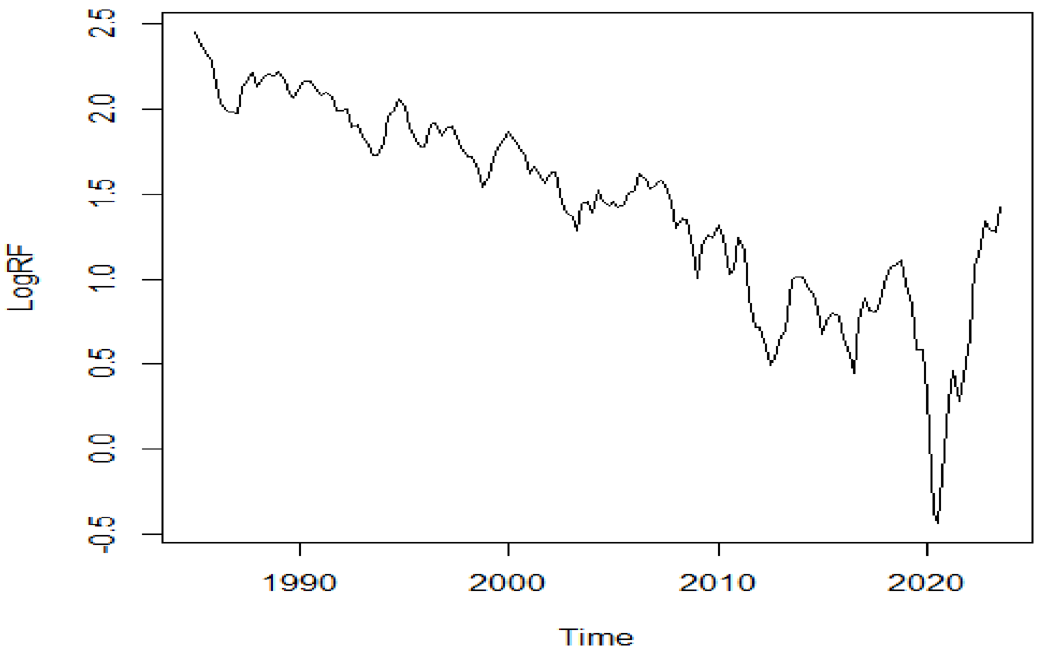
It shows a generally upward trend in the log of the S&P 500 index, indicative of long-term growth in stock market valuations. There are visible fluctuations, reflecting market volatility and various economic cycles.

Chart 2: Logarithm of US real GDP



It exhibits a consistent upward trend, suggesting steady economic growth. The slope is relatively smooth, indicating fewer short-term fluctuations compared to the stock market.

Chart 3: Logarithm of 10-year US Treasury bond yield



It appears to be in a general downtrend over the given period. The yield, which moves inversely to price, shows significant peaks and troughs, reflecting changes in interest rates and economic conditions.

ADF test to check for stationarity

First, we did an The Augmented Dickey-Fuller (ADF) test which is a **statistical test** used to determine whether a time series has a unit root and hence is not stationary. As a reminder:
H_0 : There's a unique root i.e my time serie is not stationary

To do It, we decided to use the function "ur.df" from the "urca" package for conduction the ADF test. This choice is explained by the fact that we can choose to include either a drift, a trend or neither in the test equation. Additionally, there's an automatic selection based on information criteria like the AIC which can lead to more reliable and statistically sound results.

For the S&P 500 and Real GDP, we choose **trend** in the unit root tests because both series exhibit clear long-term upward trends, reflecting consistent growth over time. Conversely, for the 10-Year Treasury Yield,

drift is selected due to its fluctuating nature around changing mean levels, without a clear long-term linear trend.

Table 1: ADF test results for the log of my series

Series	Test Type	Test Statistic	Critical Values (1%, 5%, 10%)
LogSP500	Trend	-2.259	(-3.99, -3.43, -3.13)
LogRGDP	Trend	-1.7031	(-3.99, -3.43, -3.13)
LogRF	Drift	-2.2371	(-3.46, -2.88, -2.57)

The ADF unit root tests for the log(S&P 500), log(Real GDP), and log(10-Year Treasury Yield), conducted using the Augmented Dickey-Fuller method, reveal that all three series are non-stationary. Specifically, for both the S&P 500 and Real GDP, the test statistics do not exceed the critical values at conventional significance levels, suggesting the presence of unit roots even after accounting for trends. Similarly, the log of 10-Year Treasury Yield, which was tested for stationarity around a drift, does not provide sufficient evidence to reject non-stationarity.

We also did a complementary ADF test in our code that confirmed the 3 series are not stationary. We found that p-values > 0.05 (focusing on the type 2 for logRF : With Drift No Trend).

First Difference

Now, we will see if my series are integrated of order 1. To do so, we applied the first difference to our 3 variables: logarithm of S&P500, logarithm of Real GDP and the logarithm of 10-year US treasury yield. We then applied again the same ADF test in order to show that these series are indeed I(1)

Table 2: ADF test results for differenced series

Series	Test Type	Test Statistic	Critical Values (1%, 5%, 10%)
Diff LogSP500	Trend	-8.5523	(-3.99, -3.43, -3.13)
Diff LogRGDP	Trend	-9.4613	(-3.99, -3.43, -3.13)
Diff LogRF	Drift	-8.6657	(-3.46, -2.88, -2.57)

The Augmented Dickey-Fuller unit root tests conducted on the first differences of the LogSP500, LogRGDP and LogRF demonstrate that these series are stationary as we reject the null hypothesis of the presence of a unit root at the 1% significance level. For the differenced LogSP500 and LogRGDP, the negative test statistics (-8.5523 and -9.4613, respectively) significantly exceed the critical values, indicating the rejection of the non-stationarity hypothesis. Similarly, the differenced LogRF shows a test statistic of -8.6657, which also supports stationarity. These findings suggest that the series are integrated of order 1 (I(1)), meaning they become stationary after a single differencing. This is a crucial insight for subsequent econometric modeling, such as cointegration analysis or vector error correction modeling.

Our complementary ADF test confirmed that with all p-values <= 0.01.

Now that we proved that our series are I(1), we therefore will try to find if there's a relation of cointegration among them.

Test for Cointegration

In order to check If there's indeed a relation of cointegration among my series, we are going to start by checking if there's a linear combination of them. We will regress the logarithm of S&P500 on logarithm of the Real US GDP and the logarithm of the 10-year US treasury yield. This can be represented by the following Equation :

$$\boldsymbol{log}(\boldsymbol{P_t}) = \boldsymbol{log}(\boldsymbol{D_t}) + \boldsymbol{log}(\mathbf{r^F}_t) + \boldsymbol{Z_t}$$

With Z_t as residuals that needs to be stationary in order to prove the cointegration

Table 3: Regression Summary of lm1(LogSP500 ~ LogRGDP + LogRF)

Term	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	-24.67512	1.64274	-15.021	<2e-16
LogRGDP	3.27314	0.16265	20.124	<2e-16
LogRF	0.10741	0.05809	1.849	0.0664
Residual Std. Error	0.2089			
Multiple R-squared	0.939			
Adjusted R-squared	0.9382			
F-statistic	1170			<2.2e-16

This regression table shows that LogRGDP and LogRF are statistically significant predictors of LogSP500, with positive coefficients indicating direct relationships. The intercept is negatively significant. The model has a high R-squared (0.939), suggesting it explains much of the variance in LogSP500. The residual standard error is relatively low, indicating the model's predictions are precise. The highly significant F-statistic demonstrates the overall significance of the model.

We are interested by the cointegration of my variable, so I will check if the residuals of this regression are stationary.

Table 4: ADF Test for Residuals of lm1

Parameter	Value
Test Statistic	-2.3142
Critical Value (1%)	-2.58
Critical Value (5%)	-1.95
Critical Value (10%)	-1.62
p-value	0.022

The ADF test statistic for the residuals is -2.3142, and the critical values at the 1%, 5%, and 10% levels are -2.58, -1.95, and -1.62, respectively. The test statistic is more negative than the 5% critical value but not the 1% value. With a p-value of 0.022, the null hypothesis of a unit root can be rejected at the 5% significance level, suggesting that the residuals are stationary.

This finding implies that there is evidence of a long-term equilibrium relationship (cointegration) among the log-transformed S&P 500, log of Real GDP, and the logarithm of the 10-Year Treasury Yield.

To confirm It, we tested the number of differences needed so that my residuals can be stationary and the result was ndiffs(resid) = 0.
That confirms their I(0) characteristic.

Structural Breaks

In the next phase of our analysis, we will incorporate dummy variables, denoted as $\mathbf{DUM_t}$, into the long-term equation that represents the equilibrium price of the S&P 500 index. These dummy variables are introduced to capture structural shifts in market expectations, such as changes in inflation rates and earnings growth, which are reflected by the term $\boldsymbol{\pi^e_t - g^e_t}$ in the Gordon-Shapiro model. While such shifts in expectations may be infrequent, their impact on stock price movements can be significant and enduring. To account for these shifts, we model the **DUM_t** as a constant, with the dummy variables switching to a value of 1 to signify the occurrence of a structural break, thereby capturing any prolonged alterations in the level of expected inflation and earnings growth. To be clearer, the dummies take the values of 1 after a given structural break.

First, we determine the optimal number of breaks. We used a function from the **strucchange** package in R is used to test for structural changes in linear regression models (we applied It to our previous Equation).

Table 5: Breakpoints Analysis

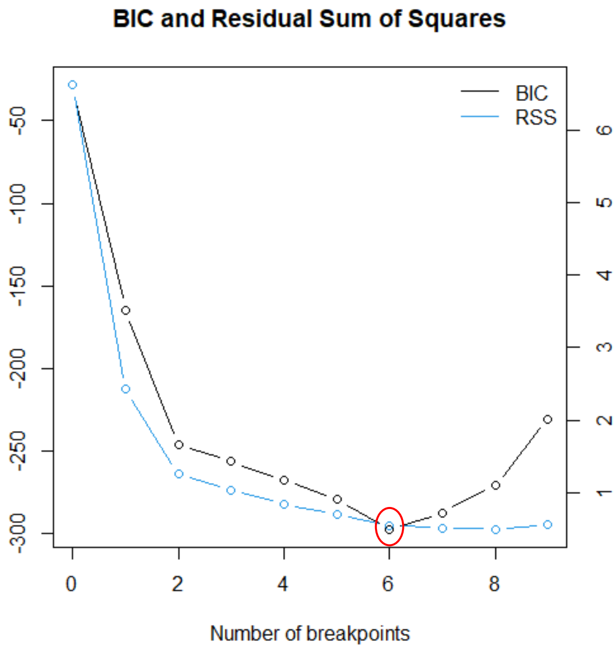
Number of Breaks (m)	Breakdates	BIC
1	2002(3)	-164.3774
2	2001(3), 2008(4)	-245.7347
3	1997(3), 2002(3), 2008(4)	-256.1944
4	1998(1), 2006(1), 2012(1), 2019(4)	-267.5623
5	1994(2), 1998(3), 2006(1), 2012(1), 2019(4)	-279.1121
6	1994(2), 1998(3), 2002(3), 2008(4), 2013(2), 2019(4)	-297.0336
7	1991(1), 1995(3), 2000(1), 2003(4), 2008(4), 2013(2), 2019(4)	-287.2770

The breakpoint analysis of the S&P 500, in relation to Real GDP and 10-Year Treasury Yield, unveils six critical structural shifts within the study period. The model with the optimal Bayesian Information Criterion (BIC) value highlights these breakpoints as occurring in early 1994, mid-1998, early 2002, late 2008, mid-2013, and late 2019 (We select the model with the lowest BIC). These periods correspond to significant economic events, including the mid-1990s market adjustments, the aftermath of the dot-com bubble burst, the onset of the global financial crisis, and the recent economic fluctuations influenced by major policy shifts and global phenomena such as the COVID-19 pandemic. Integrating these breakpoints into our econometric models offers a more nuanced understanding of how the S&P 500 index interacts with pivotal economic indicators during distinct phases of economic cycles and major global developments. This approach ensures a comprehensive analysis of the index's behavior across varying economic contexts.

The 6 breakpoints are: Q2 1994, Q3 1998, Q3 2002, Q4 2008, Q2 2013 and Q4 2019.

It's also confirmed by the below chart that there are 6 optimal breaks.

Chart 4: Optimal number of breakpoints



Now we can use this second Regression:

$$\log(P_t) = \log(D_t) - \log\left(r^F{}_t + \Pi^e{}_t - g^e{}_t\right) + Z_t$$

$\pi^e{}_t$ - $g^e{}_t$ is accounted for with the structural breaks in expectations dummy **DUM_t** variables.

Table 6: Regression Summary of lm(LogSP500 ~ LogRGDP + LogRF + break-factor(Breaks, breaks = 6))

Term	Estimate	Std. Error	t value	Pr(> |t|)
(Intercept)	-27.74365	1.33118	-20.841	< 2e−16
LogRGDP	3.65907	0.14412	25.389	< 2e−16
LogRF	-0.06140	0.03052	-2.012	0.046076
breakfactor(Breaks, breaks = 6)segment2	0.01640	0.04048	0.405	0.685969
breakfactor(Breaks, breaks = 6)segment3	0.05381	0.05966	0.902	0.368606
breakfactor(Breaks, breaks = 6)segment4	-0.49862	0.07613	-6.549	9.24e−10
breakfactor(Breaks, breaks = 6)segment5	-0.79620	0.08844	-9.003	1.09e−15
breakfactor(Breaks, breaks = 6)segment6	-0.60588	0.10505	-5.767	4.63e−8
breakfactor(Breaks, breaks = 6)segment7	-0.44055	0.12121	-3.635	0.000385
Residual Std. Error		0.09341 (df = 146)		
Multiple R-squared		0.9883		
Adjusted R-squared		0.9876		
F-statistic		1540 on 8 and 146 DF		< 2.2e−16

Table 7: ADF test for residuals of lm2

Parameter	Value	Conclusion
Test Statistic	-5.1949	Stationary
Critical Value (1%)	-2.58	
Critical Value (5%)	-1.95	
Critical Value (10%)	-1.62	
p-value	6.53×10^{-7}	

In the analysis of the linear regression model **lm2**, which examines the relationship between LogSP500, LogRGDP, LogRF, and identified breakpoints, the Augmented Dickey-Fuller test on the residuals indicates stationarity. The test statistic of -5.1949, significantly more negative than the critical values at 1%, 5%, and 10% levels, coupled with a very low p-value, confirms the absence of any inherent trends or patterns in the residuals over time. This finding is crucial as it validates the model's specification and implies that the relationships between the variables are stable and consistent throughout the period studied. In conclusion, there's indeed a relation of cointegration between the S&P500 and Its fundamentals justifying the existence of a long-term equilibrium to describe the equilibrium price with 6 structural breaks.

We will now confirm this result with a Trace test.

First, we determine the optimal number of lags following the AIC and FPE. **We found that it is 7.**

Table 8: Johansen Cointegration Test Results (Trace Test)

$r \leq$	Test Statistic	10%	5%	1%
0	76.42	49.65	53.12	60.16
1	45.35	32.00	34.91	41.07
2	15.68	17.85	19.96	24.60
3	4.10	7.52	9.24	12.97

Table 9: Johansen Cointegration Test Results (Maximal Eigenvalue Test)

$r \leq$	Test Statistic	10%	5%	1%
0	31.07	25.56	28.14	33.24
1	29.67	19.77	22.00	26.81
2	11.58	13.75	15.67	20.20
3	4.10	7.52	9.24	12.97

The Johansen Trace test and Maximal Eigenvalue test were conducted to ascertain the number of cointegrating relationships among the variables under study. The Trace test results demonstrate that we can reject the null hypothesis of no cointegration (r=0) and the hypothesis of at most one cointegrating relationship ($r \leq 1$), as the test statistics for both are greater than the critical values at the 10%, 5%, and 1% significance levels. This suggests the existence of at least two cointegrating relationships. As we can see in table 8, out test statistic for $r \leq 2$ is lower than all critical values so we can conclude that there exist at most 2 cointegration relationships.

The Maximal Eigenvalue test corroborates these findings, reinforcing the conclusion that there are two significant long-term, stable relationships among the series. These cointegrating relationships are essential for the construction of a Vector Error Correction Model (VECM), providing a foundation for understanding the equilibrium dynamics of the system.

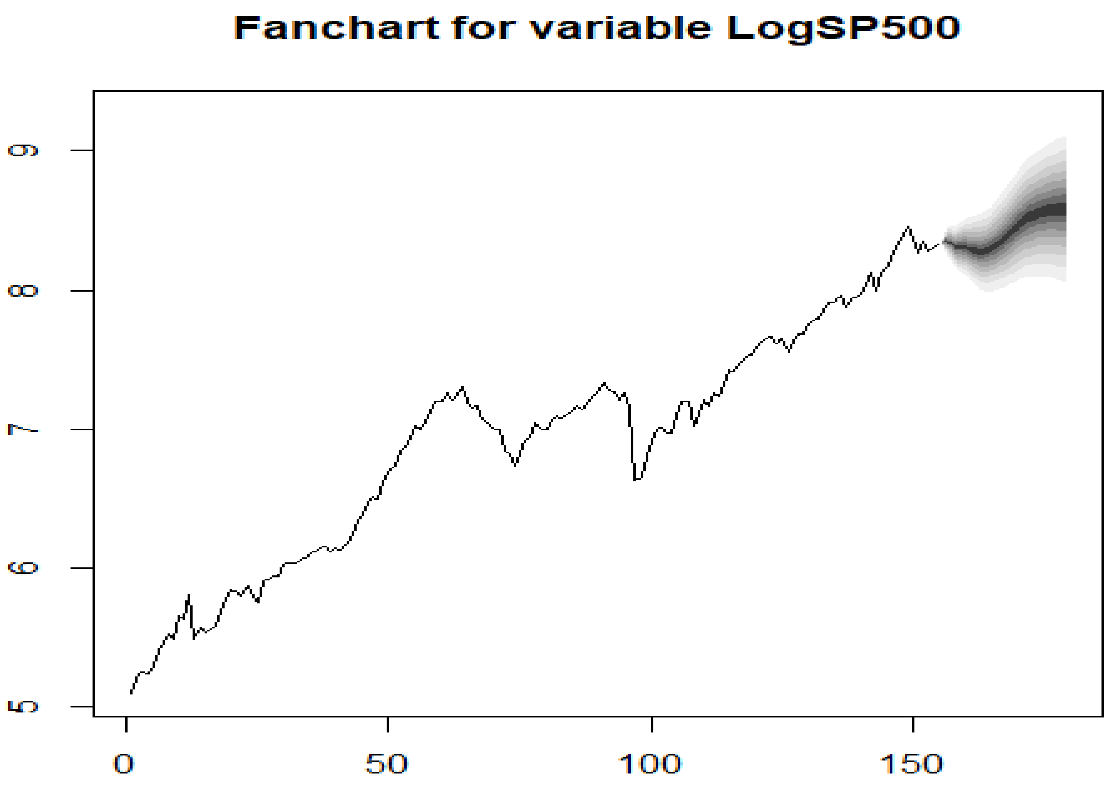
VECM

In the study, a Vector Error Correction Model (VECM) has been utilized to delve into the long-term equilibrium dynamics of three key financial series: the logarithms of the S&P 500 index, Real GDP, and the 10-Year Treasury Yield. Structural breaks within the data, identified earlier, are also incorporated as exogenous factors to improve model accuracy.

The VECM is chosen due to the presence of cointegrating relationships among the variables, suggesting that they move together in the long run, even if they might deviate in the short term. The **lag=6 (as we found lag = 7 with VAR)** parameter in the VECM function indicates that the model considers six lagged observations for each variable to predict the current value, which is typical for data with potential seasonal and inertia effects. The parameter **r=2** is based on the result of the Johansen cointegration test, which indicated two cointegrating relationships in our system.

The chart 5 below is an illustration of the forecast generated from the Vector Error Correction Model (VECM) for the logarithm of the S&P 500 index. The chart demonstrates the projected path of the index over the forecast horizon, along with the associated uncertainty depicted as shaded areas around the prediction. These shaded areas represent confidence intervals, with darker shades indicating higher levels of confidence. The chart shows the expected future values of the LogSP500, taking into account the long-term equilibrium relationship with Log Real GDP and the Log 10-Year Treasury Yield, as well as accounting for structural breaks identified in the data. The upward trend suggests an anticipation of growth in the index value, while the widening confidence bands reflect increasing uncertainty as the forecast extends further into the future.

Chart 5: Forecast of the logarithm of the S&P500 stock index 12 quarters ahead



Conclusion

- The chosen period and frequency for analyzing the (log-) prices of the stock index are selected based on historical financial events and data availability, ensuring a comprehensive representation of the index's performance.
- All series were confirmed to be I(1) after testing, which indicates that they are integrated of order one, and any trend or seasonality has been removed to ensure stationarity.
- The Gordon-Shapiro model provides a theoretical foundation for the existence of a long-term equilibrium price equation, which has been empirically validated by estimating and testing the equation, including the identification of structural breaks.
- Utilizing the Vector Error Correction Model (VECM), the study concludes a long-term equilibrium relationship among the stock index, Real GDP, and 10-Year Treasury Yield, with forecasts indicating an upward trend in the stock index amid increasing future uncertainty.