Thẻ 1

FINANCIAL ECONOMETRICS COURSEWORK

Nguyen Hoai Linh (ID: 279996)
Do Thi Huong (ID: 279995)

Academic Year: 2025/2026

Introduction

This research focuses on the investigation and application of modern financial econometric methods to address two important problems: testing for structural changes in variance and measuring nonlinear information transmission among economic variables. The main objective is to combine theoretical analysis, Monte Carlo simulation, and empirical applications using real data in order to gain a deeper understanding of the complex dynamics of financial markets.

In the first part (Part A), the research concentrates on testing for heteroscedasticity when the timing of structural breaks is unknown. Building on the Sup Goldfeld–Quandt (Sup-GQ) test, which focuses solely on the split point that yields the maximum test statistic, this study proposes a new approach that aggregates information from all potential split points through a combined p-value statistic. The performance of the proposed method is rigorously evaluated using Monte Carlo experiments, with particular emphasis on comparing its power against traditional tests under finite-sample conditions.

The second part (Part B) conducts an empirical analysis within the Quantile Transfer Entropy (QTE) framework. This part aims to investigate the directional information flow and causal effects from economic uncertainty indicators to financial market returns. Rather than focusing only on the mean, the QTE approach allows the examination of information transmission across different quantiles, thereby shedding light on whether economic uncertainty has heterogeneous effects during extreme market conditions (such as severe losses or exceptionally high returns) compared to more stable periods.

From a technical perspective, the research emphasizes the use of advanced resampling techniques due to the non-standard distributional properties of the test statistics. Specifically, the Wild Bootstrap is applied to the variance testing problem in Part A, while the Stationary Block Bootstrap is employed in Part B to preserve the time-dependence structure of the time series data. All procedures, including data processing, simulation, and hypothesis testing, are implemented in the R programming language, ensuring scientific rigor and reproducibility.

Part A: Theoretical Analysis and Power Assessment

1. **Heteroscedasticity and Homoscedasticity**

The Ordinary Least Squares (OLS) estimator is widely used in econometrics due to its simplicity and strong theoretical properties. However, when we apply the OLS method in practice, we can observe errors between the dependent variables in the estimation. To prevent the variance of the disturbances from affecting the estimation results, five important assumptions about the disturbance term have been established:

(1) $E(u_t) = 0$

(2) $var(u_t) = \sigma^2 < \infty$

(3) $cov(u_i, u_j) = 0$

(4) $cov(u_t, x_t) = 0$

(5) $u_t \sim N(0, \sigma^2)$

The second assumption states that the variance of the errors is constant, a condition known as homoscedasticity. When this assumption is violated, it means that the variance of the error term is not constant across observations. We can then conclude that there is heteroscedasticity in the regression.

Detecting heteroscedasticity in econometrics is extremely important. When heteroscedasticity occurs, the standard errors of the regression become incorrect. As a result, the variance of the coefficients(Var(β^)), t-tests, F-tests, and confidence intervals are all affected, which makes the statistical significance of the variables unreliable.

2. **Detection of Heteroscedasticity**

To detect heteroscedasticity, we can use some tests: the Goldfeld–Quandt test and the Sup-Goldfeld–Quandt test, Fisher Method, Breusch-Pagan test and White test.

2.1 Goldfeld–Quandt test and the Sup-Goldfeld–Quandt test

Both methods are based on splitting the given dataset into two smaller samples. Based on the separated samples, the two residual variances are calculated

$$s_1^2 = \frac{\hat{u}_1' \hat{u}_1}{T_1 - k} \quad \text{and} \quad s_2^2 = \frac{\hat{u}_2' \hat{u}_2}{T_2 - k}$$

Then, we perform the test with the null hypothesis that the variance of the disturbances is equal.

$$H_0: \sigma_1^2 = \sigma_2^2$$

(H0: homoscedasticity

H1: heteroscedasticity)

H_0: The variance of the error term is constant,

$$\mathrm{Var}(u_t) = \sigma^2.$$

H_1: The variance changes at an unknown break point.

The main difference between the Goldfeld–Quandt (GQ) test and the Sup-Goldfeld–Quandt (Sup-GQ) test lies in determining the splitting point. For the GQ test, we need to calculate and carefully select the exact break point to avoid errors. In reality, it is very difficult to identify the exact break point, so we cannot determine where the heteroscedasticity actually changes, or the results are unreliable when using the GQ test. In contrast, the Sup-GQ test scans all possible points within the sample and identifies the largest critical value. Consequently, it overcomes the limitations of the GQ test.

The Sup-Goldfeld–Quandt test is a powerful test for heteroscedasticity in regression models when the break point is unknown. By examining all possible split points and taking the largest test statistic, this method avoids arbitrary dependence on the researcher's choice. However, this very feature makes the test statistic's distribution non-standard, and it does not follow an F-distribution. Therefore, theoretical critical values cannot be applied. To overcome this, Wild Bootstrap or Monte Carlo simulations are used to estimate the empirical distribution and produce a test with good performance, stable size, and high power.

Sup_GQ test

In this research, we conducted the Sup-GQ test to investigate the Heteckesdascity error and the testing steps:

- Hypothesis:

  Ho: The variance of error term is constant.

  Ha: the variance of error term changes at an unknown break point.

- Build regression model $y_t = 1 + x_t + \varepsilon_t, \quad x_t \sim N(0,1)$ base on

- Specify the set of admissible break points(trimming)
- Calculate the GQ test statistic:

$$\gamma = \frac{RSS_2/_{df}}{RSS_1/_{df}}$$

- Construct the Sup-GQ test statistic:

$$F_{\text{Sup-GQ}} = \sup_{\tau \in \mathcal{T}} F_{GQ}(\tau).$$

Find out the critical value: To determine the critical value for the specified level of significance (alpha), use the F Table. The values of df1 and df2 in this test are identical (df1=df2).

Decision rule: Compare F-statistic and F-critical

```
If F_calculated < F_critical  ; Accept the Null Hypothesis.
If F_calculated > F_critical  ; Reject the Null Hypothesis.
```

In order to run the GQ and Sup-GQ tests in this assignment, we first used R to create random data. We used the function **set.seed (2025)** to fix the random number generator's initial state. The variable X was then constructed using a normal distribution. We generated a variable for the variance in R called **var_structure** in order to compute the error variance. Finally, we used a simple linear regression function to generate the variable Y by **lm() function**.

After constructing the variables **X** and **Y**, we defined the function **calc_statistics()**, which takes the dependent and independent variables as inputs and applies a trimming proportion of 15% at both the beginning and the end of the data series. Next, we constructed the set of candidate break points, denoted by **tau_grid**. To store the values of the F-statistics and the corresponding p-values at each break point, we created two new variables, **f_stats** and **p_vals.**

To calculate the F-statistics of Goldfeld–Quandt test:, We based our analysis on the formula used to compute the GQ test:

$$\gamma = \frac{RSS_2/_{df}}{RSS_1/_{df}}$$

We then proceeded to compute the values for each subsample and store the values of the F-statistics and the corresponding p-values at each break point.

Lastly, we compile the Sup-GQ test's final findings. According to the theoretical formulation, the greatest F-statistic from the Goldfeld–Quandt (GQ) tests across all candidate break points is F-statistic of Sup-GQ.

2.2 Proposing a Combined P-Value Alternative: Fisher Method

Proposal

Let τ = {τ1,τ2,…,τM} denote a grid of admissible split points after trimming the sample at both ends. For each candidate split point τm, a Goldfeld–Quandt (GQ) test is conducted, and the corresponding asymptotic p-value pm is obtained. To test the joint null hypothesis of homoscedasticity across all potential break points, we propose a combined p-value test based on Fisher's method. The combined test statistic is defined as

$$G = -2\sum_{m=1}^{M} \log(p_m).$$

Large values of G provide evidence against the null hypothesis of homoscedasticity.

Justification

The Sup-GQ test focuses on the maximum GQ statistic among all candidate break points and is effective only when a single strong variance break exists. When variance changes are mild or distributed, or when the sample size is small, the test can be sensitive to outliers and produce unstable results. Due to these limitations, the GQ test is non-standard and can occasionally produce unreliable results.

The choice of Fisher's Method over the Sup GQ test in this model is fully statistically justified. Fisher's Method for combining p-values addresses the limitations of the Sup GQ test by aggregating all p-values from independent tests rather than relying solely on the maximum value. This approach leverages the full evidence from multiple tests, enhancing overall statistical power instead of relying on a single extreme statistic like the Sup GQ. Moreover, the Sup GQ test is suitable primarily when the focus is on heteroscedasticity with an unspecified breakpoint. In contrast, Fisher's Method is more flexible since it does not require the model to adhere to a specific break structure. Therefore, Fisher's Method provides more stable, generalizable, and reliable conclusions in the context of this study.

Critical Values

Under the assumption of independent p-values, Fisher's combined statistic follows a chi-square distribution. However, this assumption is violated in the present context because the individual p-values are computed from overlapping subsamples of the same data and from closely spaced candidate split points. As a result, the p-values are dependent, and the asymptotic distribution of the combined statistic G is non-standard. Consequently, analytical critical values based on the chi-square distribution are invalid in this setting. To obtain valid inference, critical values for the proposed combined p-value test must be computed using a simulation-based approach, such as a wild bootstrap or Monte Carlo procedure. This ensures that the dependence structure among the p-values is properly accounted for and that the test maintains the correct size.

2.3. Existing Tests

a. Breusch-Pagan Test

The Breusch-Pagan test is a widely used method for detecting heteroscedasticity in econometrics. Unlike the Goldfeld-Quandt (GQ) test, it does not require ordering or splitting the dataset. While the GQ test only examines whether the variances differ between two groups of observations, the Breusch-Pagan test checks whether the variance of the errors depends linearly on one or more explanatory variables. The testing procedure involves three steps. First, the original model is estimated using OLS to obtain the residuals (u^t). Next, an auxiliary regression is performed, with the squared residuals as the dependent variable. The general form of this auxiliary regression is:

$$\hat{u}_t^2 = \alpha_1 + \alpha_2 z_{2t} + \alpha_3 z_{3t} + \cdots + \alpha_m z_{mt} + v_t$$

While z can be either the original explanatory variables x from the model or other variables suspected of causing heteroscedasticity.

Finally, the test statistic LM is calculated as LM = T × R^2, where T is the number of observations and R^2 is the coefficient of determination from the auxiliary regression. If the calculated LM value exceeds the critical value from the chi-squared distribution (or the p-value is smaller than the significance level), we reject H0 of homoscedasticity (all coefficients α in the auxiliary regression are zero, except for α1) and conclude that the model exhibits heteroscedasticity.

The Breusch-Pagan test is flexible, does not require splitting the sample, can test multiple variables simultaneously, and serves as the foundation for the White test, which extends the method to detect nonlinear variance changes and interactions among variables.

b. White Test

Besides the three tests above, the White test can also be applied to check for heteroscedasticity. This test applies to both linear and nonlinear models. To perform this test, the first step is to estimate the original model, and then we can obtain the residuals. Next, an auxiliary regression is constructed using these residuals. The unrestricted model allows the squared residuals to be regressed on a constant and the original explanatory variables. This regression is considered the unrestricted model.

$$\boldsymbol{\mu_i^2} = \boldsymbol{\alpha_1} + \boldsymbol{\alpha_2 X_{1i}} + \boldsymbol{\alpha_3 X_{2i}} + \boldsymbol{\alpha_4 X_{3i}} + \boldsymbol{\alpha_5 X_{1i}^2} + \boldsymbol{\alpha_6 X_{2i}^2} + \boldsymbol{\alpha_7 X_{3i}^2} + \boldsymbol{\alpha_8 X_{1i} X_{2i}} + \boldsymbol{\alpha_9 X_{1i} X_{3i}} + \boldsymbol{\alpha_{10} X_{2i} X_{3i}} + \boldsymbol{v_i}$$

To test the hypothesis, we can rely on the F-statistic. We need to estimate an additional restricted model, in which the squared residuals are regressed only on a constant, corresponding to the null hypothesis that the variance of the errors is constant. Finally, the total sum of squares of the residuals from these two models is used to compute the RSS in the F-statistic formula.

Alternatively, the hypothesis can also be tested using the LM statistic, with the following formula:

$$White\ Statistic = n \cdot R^2 \sim \chi^2_{df}$$

Here, n is the number of observations, R-squared comes from the unrestricted regression model, and df is the number of explanatory variables in the unrestricted regression. Finally, the test statistic is compared to the chi-square distribution to draw a conclusion about the null hypothesis.

Although the White test does not require the assumption of normally distributed errors and can be easily applied without selecting a break point, its results may be less reliable for small sample sizes.

3. Monte Carlo power comparison

In Part A of this project, the Monte Carlo simulation is used to evaluate the finite-sample size and power of the Sup-GQ test and the proposed combined p-value test. Since both test statistics have non-standard distributions or no known distribution and rely on bootstrap critical values, analytical results are not available. Monte Carlo allows us to repeatedly generate controlled datasets, apply each test, and estimate empirical rejection frequencies, providing a practical assessment of their performance. In other words, Monte Carlo simulation estimates a value by running many random scenarios and observing the outcomes, rather than solving the problem analytically. The more simulations we do, the closer the average outcome gets to the true value.

We create the data via the Monte Carlo simulation based on the following model:

$$y_t = 1 + x_t + \varepsilon_t, \quad x_t \sim N(0,1)$$

While the error variance should be defined as:

$$\epsilon_t \sim N(0, \sigma_t^2) \quad \text{where} \quad \sigma_t^2 = 1 + \delta \cdot I(t/N > 0.5)$$

We set the variance break at 0.5 of the sample. This produces two equal-sized subsamples, which ensures stable estimation of the variances on both sides and avoids boundary bias. It also represents a standard benchmark scenario for evaluating heteroscedasticity tests. δ controls the magnitude of the variance shift. When δ = 0, we are under the null and evaluate empirical size. When δ = 1 and δ = 3, we generate moderate and strong heteroscedasticity, respectively, allowing us to study how quickly each test gains power as the variance break becomes more pronounced. We use a sample size of N = 100 for a balance between reliability and realism, allowing 15% trimming. With R = 1000 Monte Carlo replications, the results are statistically reliable, minimizing Monte Carlo noise and enabling meaningful comparison of empirical size and power across tests.

4. Empirical Results

Table 1: Empirical Size (δ=0) and Power (δ=1,3) at 5% Significance Level

Delta	Sup-GQ	Fisher's G	Breusch-Pagan	White
0	0.058	0.046	0.048	0.041
1	0.413	0.690	0.041	0.049
3	0.974	0.998	0.051	0.045

Interpretation:
- Delta = 0: Size check (should be ≈ 0.05 = nominal level)
- Delta = 1: Variance doubles after midpoint
- Delta = 3: Variance quadruples after midpoint
- Higher power = better at detecting heteroscedasticity

Based on the result table from R-Studio, we can observe the following:

With delta = 0, meaning that the variance of the model does not change, there is no heteroscedasticity in the model. The detection rates for the four tests—Sup-GQ, Fisher method, Breusch-Pagan, and White test—are 5.8%, 4.6%, 4.8%, and 4.1%, respectively, indicating that all four tests perform well at the significance level.

With delta = 1, when the model variance doubles after the breakpoint, the power of Sup-GQ is 0.413, meaning it detects heteroscedasticity about 41.3% of the time. The Fisher method shows a power of 0.69, detecting about 69% of heteroscedasticity cases. In contrast, the Breusch-Pagan test has a power of only 0.041 (4.1%), and the White test detects heteroscedasticity at about 4.5%.

With delta = 3, when the model variance quadruples after the breakpoint, the power of Sup-GQ reaches 0.974, showing a very high ability to detect heteroscedasticity (97.4%). The Fisher test is even stronger, with an almost perfect detection rate of 99.8%. Meanwhile, the power of Breusch-Pagan remains low at 5.1%, and the White test also has a very low detection rate of about 4.5%.

From these results, it is clear that when the variance change is small, all four tests perform similarly. However, when the variance change is large, Sup-GQ and Fisher method have significantly higher detection rates than the other two tests. This difference likely arises from how each test calculates its critical value: Sup-GQ and Fisher method rely on changes in variance across positions in the sample, whereas Breusch-Pagan and White tests mainly rely on the explanatory variable x. Regardless of the magnitude of the variance change, the Fisher method remains the most effective test for detecting heteroscedasticity, demonstrating the optimality discussed above.

PART B: Empirical Application - Quantile Transfer Entropy (QTE)

2.1. Background: Uncertainty and Financial Returns

To apply Quantile Transfer Entropy (QTE) in the field of economics, we use two datasets: S&P 500 stock data from Yahoo Finance and Economic Policy Uncertainty (EPU) data for the U.S. market, covering the period from January 1, 2015 to December 31, 2025. In this study, we examine whether S&P 500 prices are affected by EPU.

In the first step, S&P 500 stock data from Yahoo Finance are imported into R using the getSymbols() function. After that, the write.csv command is used to save the S&P 500 data as an Excel file.
Next, the data are processed and a new Excel file is created, which includes closing prices and EPU indicators from January 1, 2015 to December 31, 2025.
Finally, the dataset is imported into R using the **read.csv()** command with the file name "EPU_SP.csv".

To conduct the QTE test, the S&P 500 index is transformed into log returns by diff(log()) function, and all NA (missing) values in the new return data are removed.

2.2. The QTE Framework

a. Definition

Quantile Transfer Entropy (QTE) is an extension of Transfer Entropy designed to measure the magnitude and direction of information transmission from process X to process Y. Still, instead of evaluating the entire distribution as in the standard TE framework, QTE focuses on specific quantiles of the distribution of Y. Specifically, the $QTE_{\tau,k}$ statistic compares the conditional entropy of Y_t given its own past with the conditional entropy of Y_t when both the past of Y and the past of X are taken into account. If the inclusion of past information from X leads to a meaningful reduction in the uncertainty of Y_t at quantile τ, this implies that X provides additional predictive information for Y at that distributional level. Therefore, instead of merely answering whether "X affects Y," QTE allows us to determine under which conditions such an effect occurs. For example, during normal periods, severe downturns (lower tail), or strong expansion phases (upper tail). In particular, QTE measures the directional information flow from the uncertainty index to stock returns at different parts of the return distribution. For instance:

- $\tau = 0.1$ (lower quantile): corresponds to extreme losses, indicating that uncertainty has a crucial predictive role during market downturns.
- $\tau = 0.5$ (median): reflects typical market conditions.
- $\tau = 0.9$ (upper quantile): corresponds to extreme gains, highlighting predictive effects during strong market upturns.

From a statistical perspective, $QTE_{\tau,k}$ is non-negative, and values significantly greater than zero indicate the presence of information transfer from X to Y at quantile τ. To formally assess this, we conduct the following hypothesis test:

H0: $QTE_{\tau,k} = 0$ (no information transfer from X to Y)

Ha: $QTE_{\tau,k} > 0$ (information transfer from X to Y)

To accurately assess the information flow, QTE is based on comparing the prediction errors from two quantile regression models:

- Restricted Model: Only the lagged value of y_{t-k} is used to predict the τ quantile of y_t

$$Q_\tau(y_t|y_{t-k}) = \alpha_0 + \alpha_1 y_{t-k}.$$

- Unrestricted Model: predicts the τ quantile of y_t using both its own lag and the lag of the predictor x_{t-k}

$$Q_\tau(y_t|y_{t-k}, x_{t-k}) = \beta_0 + \beta_1 y_{t-k} + \beta_2 x_{t-k}$$

The QTE statistic is then calculated using the check function

$$\rho_\tau(u) = u \cdot (\tau - I(u < 0))$$

to evaluate the residuals of each model:

$$QTE_\tau = \log\left(\frac{\sum_t \rho_\tau(\hat{u}_{1,t})}{\sum_t \rho_\tau(\hat{u}_{2,t})}\right)$$

where and u1, u2 represent the residuals (prediction errors) from the restricted and unrestricted models, respectively. QTE measures the extra information provided by the unrestricted model at quantile τ

However, the sampling distribution of $QTE_{\tau,k}$ under H0 is non-standard, meaning that conventional critical values are not applicable. Therefore, a bootstrap procedure is implemented to approximate the null distribution and provide valid statistical inference regarding the significance of the QTE estimates.

Before applying the QTE test, the stationarity of the time series must be checked. QTE measures information flow based on probability distributions. If the data are non-stationary, the detected information transfer may be misleading and driven by trends rather than real interactions. Therefore, ensuring stationarity helps QTE capture true information transmission across different quantiles. In this study, the Augmented Dickey–Fuller (ADF) test is employed to examine the stationarity of the time series.

b. Stationarity Testing with the Augmented Dickey–Fuller (ADF) Test

Before conducting the QTE analysis, we test the stationarity of the dataset using the Augmented Dickey–Fuller (ADF) test.
The ADF test is performed by using the **adf.test()** function in R.

With the null hypothesis H_0, the data are non-stationary, while under the alternative hypothesis H_1, the data are stationary.
We reject H_0 based on the p-value. If the p-value from the test is smaller than 0.05, we reject H_0 and conclude that the data are stationary.

When we perform the ADF test on the S&P 500 return series, we obtain the following results from the R output:

```
        Augmented Dickey-Fuller Test

data:  returns
Dickey-Fuller = -5.105, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

With a lag length of 4, the p-value is 0.01, which is smaller than 0.05, so we reject H_0. Therefore, we can conclude that the S&P 500 return series is stationary.

When we perform the ADF test on the EPU index, we obtain the following results from the R output:

```
        Augmented Dickey-Fuller Test

data:  data$EPU
Dickey-Fuller = -2.7249, Lag order = 4, p-value = 0.2755
alternative hypothesis: stationary
```

With a lag length of 4, the p-value is 0.2755, which is greater than 0.05, so we do not reject H_0.
Therefore, the EPU index is not stationary.

To satisfy the assumptions required for the QTE analysis, we take the first difference of the return series and the EPU index.
Finally, all NA (missing) values in the dataset are removed.

c. Quantile Regression (Constructing restricted and unrestricted models)

To examine whether past returns (Return_lag) affect current returns (Return), and to assess whether economic policy uncertainty (dEPU_lag) provides additional predictive information for returns, we construct two quantile regression models, namely a restricted and an unrestricted specification, across multiple quantiles. The restricted model includes only lagged returns to capture the intrinsic dynamics of the return process, while the unrestricted model adds this specification by incorporating lagged changes in economic policy uncertainty in order to evaluate the marginal informational contribution of policy uncertainty. By estimating the models at different quantiles, we can see how the effects of these variables change under different market conditions. Specifically, lower quantiles represent periods of market downturns or adverse conditions, the median quantile reflects normal market states, whereas higher quantiles correspond to periods of market expansions or favorable conditions.

Restricted Model

We construct the restricted model in RStudio using the quantile regression function rq() to estimate the relationship at the 25th quantile of the dependent variable Return and obtain the following results:

Variable	Coefficient	Lower Bound	Upper Bound
Intercept	-0.00803	-0.01188	0.00012
Return_lag	0.25220	0.04909	0.36130

Examining the estimation results at the 25th quantile, the intercept is −0.00803 with a confidence interval of [−0.01188, 0.00012], which includes zero, indicating that it is not statistically different from zero. In contrast, the coefficient of Return_lag is 0.25220 with a confidence interval of [0.04909; 0.36130], entirely positive, indicating that lagged returns have a positive and statistically significant effect on the 25th quantile of current returns. That means, when the market is slightly declining, Return exhibits autocorrelation.

Unrestricted Model

Next, we construct the unrestricted model in RStudio, including both Return_lag and dEPU_lag, and obtain the following results:

Variable	Coefficient	Lower Bound	Upper Bound
Intercept	-0.02215	-0.03915	0.00688
Return_lag	0.18094	-0.05283	0.34710
DEPU_lag	0.00012	-0.00015	0.00024

The results show that the intercept is −0.02215 with a confidence interval of [−0.03915; 0.00688], which includes zero and is therefore not statistically significant. The coefficient on Return_lag is 0.18094, positive in sign but with a confidence interval of [−0.05283; 0.34710] that also includes zero, indicating that the effect of lagged returns is no longer statistically significant once dEPU_lag is added to the model. Similarly, the coefficient on dEPU_lag is very small (0.00012) and not statistically significant. This suggests that under mildly declining market conditions (the 25th quantile), economic policy uncertainty does not play an important role in explaining return dynamics, and the evidence of return autocorrelation becomes weaker compared to the restricted model.

However, the true distribution of the variable Return is unknown. Therefore, we estimate the probability density of Return. Specifically, the density at quantile $\tau = 0.25$ is estimated using the following formula:

$$f(Q(\tau)) \approx \frac{Q(\tau + h) - Q(\tau - h)}{2h}$$

where $Q(\tau)$ is the quantile function of Return, calculated using a finite difference of $h=0.05$. The estimated value of 0.0866 indicates that the density of Return around the 25th percentile is relatively low.

```
> print(f_tau_check)
      30%
0.0866284
```

d. Stationary Block Bootstrap

In this study, we conducted a statistical significance test of the QTE using the stationary block bootstrapping method in order to generate multiple simulated datasets from the original data. This allows us to observe the random distribution of the QTE and calculate the p-value of the test. Based on the comparison of the p-value with the significance level ($\alpha = 0.05$), we can make the final decision for the test.

In this study, we performed bootstrap with 499 iterations and the quantiles tested (τ) were 0.1, 0.5, and 0.9, with a lag k = 1 month. We then implemented a quick QTE calculation function to estimate the variability of the QTE efficiently. Finally, we ran the bootstrap and stored the results in the matrix *QTE_boot*.

Finally, we calculated the p-value. For each quantile, we counted the number of times the QTE values from the bootstrap were greater than or equal to the QTE from the original dataset, and we used the following formula to calculate the p-value.

$$\text{P-value}(\tau, k) = \frac{1}{B}\sum_{b=1}^{B} I\left(\widehat{Z}^{*b} \geq \widehat{\text{QTE}}(\tau, k)\right)$$

where: B is Number of bootstrap replications

$QTE\tau, k$: estimated Quantile Transfer Entropy computed from the original data

Z*b :QTE statistics obtained from bootstrap or permutation samples

2.3. Empirical Result

a. *Hypothesis testing:*

- Hypothesis:

H0: QTEτ,k = 0 (no information transfer from EPU to Return)

With τ =0.1, H0: no information transfer from EPU to 10% of low return.

With τ =0.5, H0: no information transfer from EPU to the median of returns.

With τ =0.9, H0: no information transfer from EPU to the 90th percentile of returns.

Ha: QTEτ,k > 0 (information transfer from EPU to Return)

- Test statistic:

$$QTE_\tau = \log\left(\frac{\sum_t \rho_\tau(\hat{u}_{1,t})}{\sum_t \rho_\tau(\hat{u}_{2,t})}\right)$$

- Decision rule: We reject Ho if the p-value of the testing is lower than significant level=0.05.

b. *Empirical Result :*

Quantile (τ)	State	QTE_{τ,k}	95% Bootstrap Quantile	Bootstrap p-value	Significance (α = 0.05)
0.10	Low loss	0.00393	[−0.00000 , 0.05814]	0.582	No
0.50	Median	0.04361	[0.00000 , 0.02420]	0.004	Yes
0.90	High gain	0.13192	[−0.00000 , 0.04782]	0.002	Yes

Based on the statistical results obtained from running the model in RStudio, we observe that the QTE values are non-negative across all three quantiles, indicating that EPU has the potential to transmit information and affect financial returns. However, the magnitude and statistical significance of this effect differ across quantile levels τ.

At the lower quantile $\tau=0.1$, which represents the lowest 10% of S&P 500 stock returns over the period from early 2015 to the end of 2025, the estimated QTE value is relatively small at 0.0039. The corresponding p-value is 0.582, and the 95% bootstrap confidence interval includes zero; therefore, we fail to reject the null hypothesis. This result indicates that EPU does not transmit statistically significant information to the lowest 10% of return outcomes.

In contrast, at the median quantile $\tau = 0.5$, the estimated QTE increases to 0.0436. The bootstrap p-value equals 0.004, which is well below the significance level $\alpha=0.05$, and the corresponding bootstrap confidence interval [0.00000, 0.024195] lies entirely above zero. These results provide strong evidence to reject the null hypothesis. This finding implies that Economic Policy Uncertainty (EPU) contains significant predictive information for Financial Returns under normal market conditions.

Moving to the upper quantile $\tau = 0.9$, which represents periods of large positive returns, the estimated QTE rises substantially to 0.13192. The associated bootstrap p-value is extremely small (0.002), again below the 5% significance level, indicating a strong rejection of the null hypothesis. This result suggests that EPU exerts a pronounced and statistically significant impact on returns at the highest quantile of the return distribution.

Overall, these findings reveal a nonlinear relationship between EPU and Financial Returns. While EPU does not appear to effectively signal downside risks during extreme market downturns, it plays a crucial role in influencing return dynamics under normal and strongly bullish market conditions. The marked increase in QTE values from the median quantile $\tau = 0.5$ to the upper quantile $\tau = 0.9$ indicates that information transmission from EPU is strongest during favorable market states.

From an economic perspective, EPU indices can therefore provide valuable information for both investors and policymakers when analyzing market reactions across different market regimes, rather than being viewed solely as a source of negative risk.

Conclusion

In summary, in this study, we conducted two main tests: one to detect heteroscedasticity and another to examine the information transmission of variables in the model. For the first test, we applied Monte Carlo and Wild Bootstrap methods to compare the detection rates of heteroscedasticity across four tests: Sup-Goldfeld–Quandt, Fisher, Breusch-Pagan, and White. The results indicate that the Fisher method is the most effective, as it overcomes several limitations of the other tests. Regarding the analysis of information transmission, we used Quantile Transfer Entropy with iterative Block Bootstrap to investigate the impact of Economic Policy Uncertainty (EPU) on S&P 500 stock prices from 2015 to 2025. The results show that at low quantiles, EPU has no significant effect on SP500 returns, whereas at medium and high quantiles, the influence of EPU on stock prices becomes noticeable.

REFERENCES

itfeature.com. (n.d.). *Goldfeld–Quandt test for heteroskedasticity*. Retrieved from https://itfeature.com/hetero/tests/goldfeld-quandt-test/

GeeksforGeeks. (n.d.). *Goldfeld Quandt Test in Machine Learning*. Retrieved from https://www.geeksforgeeks.org/machine-learning/goldfeld-quandt-test/

GeeksforGeeks. (n.d.). *Augmented Dickey-Fuller (ADF) Test in Machine Learning*. Retrieved from https://www.geeksforgeeks.org/machine-learning/augmented-dickey-fuller-adf/

Phân Tích Stata. (n.d.). *Phân biệt kiểm định Breusch-Pagan và White trong Stata — lựa chọn nào để phát hiện phương sai thay đổi?* Retrieved from https://phantichstata.com/phan-biet-kiem-dinh-breusch-pagan-va-white-trong-stata-lua-chon-nao-de-phat-hien-phuong-sai-thay-doi.html

SpurEconomics. (n.d.). *White test for heteroscedasticity*. Retrieved from https://spureconomics.com/white-test-for-heteroscedasticity/

LuanVanViet.com. (n.d.). *Kiểm định White trong SPSS*. Retrieved from https://luanvanviet.com/kiem-dinh-white-trong-spss/

Thẻ 2

To apply Quantile Transfer Entropy (QTE) in the field of economics, we use two datasets: S&P 500 stock data from Yahoo Finance and Economic Policy Uncertainty (EPU) data for the U.S. market, covering the period from January 1, 2015 to December 31, 2025.

In this study, we examine whether S&P 500 prices are affected by EPU.

In the first step, S&P 500 stock data from Yahoo Finance are imported into R using the getSymbols() function. After that, the write.csv command is used to save the S&P 500 data as an Excel file.

Next, the data are processed, and a new Excel file is created, which includes closing prices and EPU indicators from January 1, 2015 to December 31, 2025.

Finally, the dataset is imported into R using the read.csv command with the file name "EPU_SP.csv".

To conduct the QTE test, the S&P 500 index is transformed into log returns, and all NA (missing) values in the new return data are removed.

Stationarity Testing with the Augmented Dickey–Fuller (ADF) Test

Before conducting the QTE analysis, we test the stationarity of the dataset using the Augmented Dickey–Fuller (ADF) test.

The ADF test is performed by using the adf.test() function in R.

```
adf.test(returns)
#Dickey-Fuller = -5.105, Lag order = 4, p-value = 0.01
adf.test(data$EPU)
#Dickey-Fuller = -2.7249, Lag order = 4, p-value = 0.2755
```

With the null hypothesis H_0, the data are non-stationary, while under the alternative hypothesis H_1, the data are stationary.

We reject H_0 based on the p-value. If the p-value from the test is smaller than 0.05, we reject H_0 and conclude that the data are stationary.

When we perform the ADF test on the S&P 500 return series, we obtain the following results from the R output:

```
	Augmented Dickey-Fuller Test

data:  returns
Dickey-Fuller = -5.105, Lag order = 4, p-value = 0.01
alternative hypothesis: stationary
```

With a lag length of 4, the p-value is 0.01, which is smaller than 0.05, so we reject H_0.
Therefore, we can conclude that the S&P 500 return series is stationary.

When we perform the ADF test on the EPU index, we obtain the following results from the R output:

```
> adf.test(data$EPU)

	Augmented Dickey-Fuller Test

data:  data$EPU
Dickey-Fuller = -2.7249, Lag order = 4, p-value = 0.2755
alternative hypothesis: stationary
```

With a lag length of 4, the p-value is 0.2755, which is greater than 0.05, so we do not reject H_0.
Therefore, the EPU index is not stationary.

To satisfy the assumptions required for the QTE analysis, we take the first difference of the return series and the EPU index.
Finally, all NA (missing) values in the dataset are removed.

```
diff_returns <- diff(returns)
diff_EPU <- diff(data$EPU)
data$Return <- c(NA, diff(log(data$Close)))

data$EPU_diff <- c(NA, diff(data$EPU))
#bỏ NA
data_clean <- na.omit(data)
```

1. thêm intro của QTE+ fisher
2. Lý thuyết fisher ở phần I.
3. giải thích kết quả với delta=0, delta=1 và delta=3(phần A)- dự định thay vì chạy 1000 và 499, sẽ chạy 2-3 lần vòng lặp nhỏ hơn và kết luận,

r-300

r-100

r-500

4. giải thích cahcs làm data clean từ code 93-108(tạo ra return_lag)
5. giải thích cách xây dựng mô hình restricted và unrestricted
6. Giiar thích code tính QTE
7. Giải thích Boostrap.

4-6(L)

5-7(H)

vì sao dùng QTE QTE_exact <- f_tau * theta? thay vì công thuwccs thầy đưa ra trong đề.

We acknowledge that quantile transfer entropy based on log-loss offers a rigorous information-theoretic measure of causality. However, our primary interest lies in the economically interpretable marginal effect across different quantiles. The approximation fτ(⋅)θτf_\tau(\cdot)\theta_\taufτ(⋅)θτ serves as a local representation of quantile transfer entropy and has become standard in empirical quantile regression studies. Using this approach allows for more transparent interpretation and robust inference without altering the qualitative conclusions.

8. Giải thích kết quả từ mô hình bược 5 và kết quả từ QTE(kq bước 5 sẽ cho ta thấy được ý nghĩ thống kê của biến x lên biến Y, bược 6 giải thích biến X có ảnh hưởng lên đoạn cụ thể của Y hay không)+ giải thích kết quả của Boostrap

4. After obtaining a dataset consisting of two variables—the return of the S&P 500 index and the Economic Policy Uncertainty (EPU) index—we proceed to construct two models to predict S&P 500 returns. The dependent variable is the S&P 500 return, while the independent variables are the lagged return and the lagged first difference of EPU.

In the first model, the S&P 500 return depends solely on its lagged return.
In the second model, the S&P 500 return depends on both the lagged return and the lagged first difference of EPU.

To construct these two models, we generate lagged variables for the independent variables using the following code:

```
data_clean$Return_lag <- dplyr::lag(data_clean$Return, 1)
data_clean$dEPU_lag <- dplyr::lag(data_clean$EPU, 1)
```

Afterwards, we also examine the standard deviations of the two newly created variables in order to determine whether the generated data contain any NA values or abnormal observations. Finally, we standardize the data once again and construct a new dataset after removing all NA values.

```
# Kiểm tra độ lệch chuẩn
sd(data_clean$Return)        # Nên ~0.04 (4% hàng thá
sd(data_clean$dEPU_lag)      # ???
# Nên làm trước khi chạy mô hình
data_clean$Return_std <- scale(data_clean$Return)
data_clean$dEPU_std <- scale(data_clean$dEPU_lag)

# Loại bỏ NA do tạo độ trễ
data_clean <- na.omit(data_clean)
```

6. QTE code:

Trong kiểm định Sup GQ, giá trị thống kê sup (giá trị cực đại của các GQ statistic) thường không tuân theo phân phối chuẩn, đặc biệt khi mẫu nhỏ hoặc khi có nhiều điểm chia khả thi. Điều này làm p-value từ Sup GQ có thể không chính xác, đồng thời kết quả cũng nhạy cảm với việc loại bỏ một số quan sát (trimming) hoặc sự xuất hiện của outlier. Những hạn chế này khiến Sup GQ trở nên non-standard và đôi khi dẫn tới kết luận không ổn định trong thực tế.

Việc lựa chọn Fisher's Method thay vì Sup GQ trong mô hình này là hoàn toàn có cơ sở thống kê. Fisher's Method for Combining p-values khắc phục những hạn chế này bằng cách tổng hợp tất cả p-value từ các kiểm định độc lập, thay vì chỉ dựa vào giá trị cực đại. Fisher's Method dựa trên nền tảng lý thuyết vững chắc khi đại lượng $-2\sum \ln(p_i)$ tuân theo phân phối chi-square X^2, cho phép kết hợp nhiều p-value thành một kiểm định hợp lệ với tính nhất quán thống kê. Phương pháp này tận dụng toàn bộ bằng chứng từ nhiều kiểm định, giúp tăng sức mạnh thống kê tổng thể thay vì chỉ dựa vào một statistic cực trị như Sup GQ. Hơn nữa, Sup GQ chỉ phù hợp khi mục tiêu tập trung riêng vào heteroscedasticity với một breakpoint không xác định, trong khi Fisher linh hoạt hơn vì không yêu cầu mô hình phải tuân theo cấu trúc break cụ thể. Do đó, Fisher's Method mang lại kết luận ổn định, tổng quát và đáng tin cậy hơn trong bối cảnh nghiên cứu này.

Tại sao cần so sánh QTE từ bootstrap với QTE thực tế?

1. **QTE thực tế chỉ là một ước lượng từ dữ liệu gốc**
   Khi chúng ta tính QTE (Quantile Treatment Effect) từ dữ liệu quan sát, giá trị này chỉ phản ánh **một mẫu dữ liệu cụ thể**. Nó chưa cho biết liệu kết quả này có phải là **ngẫu nhiên** hay thực sự **có ý nghĩa thống kê**.

2. **Bootstrap tạo ra phân phối giả lập của QTE**
   Phương pháp bootstrap lấy mẫu lại nhiều lần từ dữ liệu gốc để tạo ra các **mẫu dữ liệu giả lập**. Khi tính QTE trên từng mẫu bootstrap, chúng ta sẽ có **một phân phối QTE ngẫu nhiên**. Phân phối này phản ánh **biến thiên tự nhiên** của QTE nếu dữ liệu thực tế được lấy nhiều lần.

3. **So sánh với QTE thực tế để đánh giá ý nghĩa thống kê**
   Khi đã có phân phối bootstrap, chúng ta so sánh **QTE thực tế** với **các giá trị QTE từ bootstrap**:

   - Nếu QTE thực tế nằm trong phạm vi mà hầu hết QTE bootstrap tạo ra, thì giá trị thực tế có thể là **do ngẫu nhiên** → không có ý nghĩa thống kê.

   - Nếu QTE thực tế **lớn hơn hầu hết QTE bootstrap**, thì đây là **bằng chứng mạnh mẽ** rằng hiệu ứng quan sát được không phải do ngẫu nhiên

→ có ý nghĩa thống kê.

4. **Tính p-value từ so sánh này**
   P-value được tính dựa trên **tỷ lệ số lần QTE bootstrap ≥ QTE thực tế**. Đây là cách trực tiếp để biết xác suất rằng QTE thực tế xuất hiện **do ngẫu nhiên**, từ đó đưa ra quyết định bác bỏ hay không bác bỏ giả thuyết gốc (H_0).