*Nova School of Business and Economics*

# Assignment 2

*Authors:*
Benjamin C. Herbert (45775)
Niclas Frederic Sturm (45914)
Thomas Dornigg (41727)
Vanderhulst Michel (45183)
Beatriz Vidal Rodríguez (40757)

*Supervisors:*
Prof. Paulo M. M. Rodrigues
TA. Daniel Belo

An assignment submitted for the course

*2272 Financial Econometrics*

March 14, 2021

# Part I

The following analysis is based on the data provided in the spreadsheet *PredictorData2019* from Goyal. As stated in the task, only the *Index* variable is considered for the following analysis. The monthly time series contains 1788 observations, ranging from January 1871 till December 2019.

## 1 Estimate an AR(1) model for the monthly Index log-return series and test for ARCH effects

In the following, the group will first examine the time series for stationarity by using the plot as shown in Figure 1 below. In a next step, we will test with the **A**ugmented **D**icky **F**uller (ADF) and the KPSS test to validate our analysis.

A time series process $\{Y_t \mid t \in T\}$ is weakly stationary (or second-order stationary) if

1. The mean function is constant and finite - $\mu_t = \mathbb{E}[Y_t] = \mu < \infty$
2. The variance function is constant and finite - $\sigma_t^2 = \mathrm{Var}[Y_t] = \sigma^2 < \infty$
3. The autocovariance and autocorrelation functions only depend on the lag

As stated in the task, the computation of the monthly log-returns is required to model the following processes. Log returns have some more favourable properties for statistical analysis than the simple net returns $R_t$. Mathematically speaking, they can be formulated as:

$$r_t = \ln(1 + R_t) = \ln\left(\frac{S_t}{S_{t-1}}\right) = \ln(S_t) - \ln(S_{t-1}) \tag{1}$$

As for practical meaning, academics concentrate on log returns because they eliminate the non-stationary properties of the data set, making the financial data more stable. Also, it is important to mention that log-returns are independent and identically distributed (i.i.d.).

From the index series, we take the log returns, as seen in Table 1, by applying equation (1).

| date | index | log returns |
|---|---|---|
| 1871-01 | 4.440000 | NaN |
| 1871-02 | 4.500000 | 0.013423 |
| ... | ... | ... |
| 2019-10 | 3037.560059 | 0.020226 |
| 2019-11 | 3140.979980 | 0.033480 |
| 2019-12 | 3230.780029 | 0.028189 |

Table 1: Index log returns

The starting point for the examination of the time series for stationarity is the plot in Figure 1 below. Observing the log-returns in the graph it can be assumed that the series is most likely stationary. This is also clearly observable, when looking at the rolling standard-deviation and mean in the plot. Both lines have a constant variation in the beginning, which indicates a stationary behaviour. One exception from this stationary property can be observed around the 1930s, where log-returns, rolling mean and standard deviation show an upward movement respectively outbreak which most likely has its roots in the Great Depression, starting in August 1929. However, all in all it can be said that even though small breaks occur, a first-order differencing may not be required to transform the time-series.

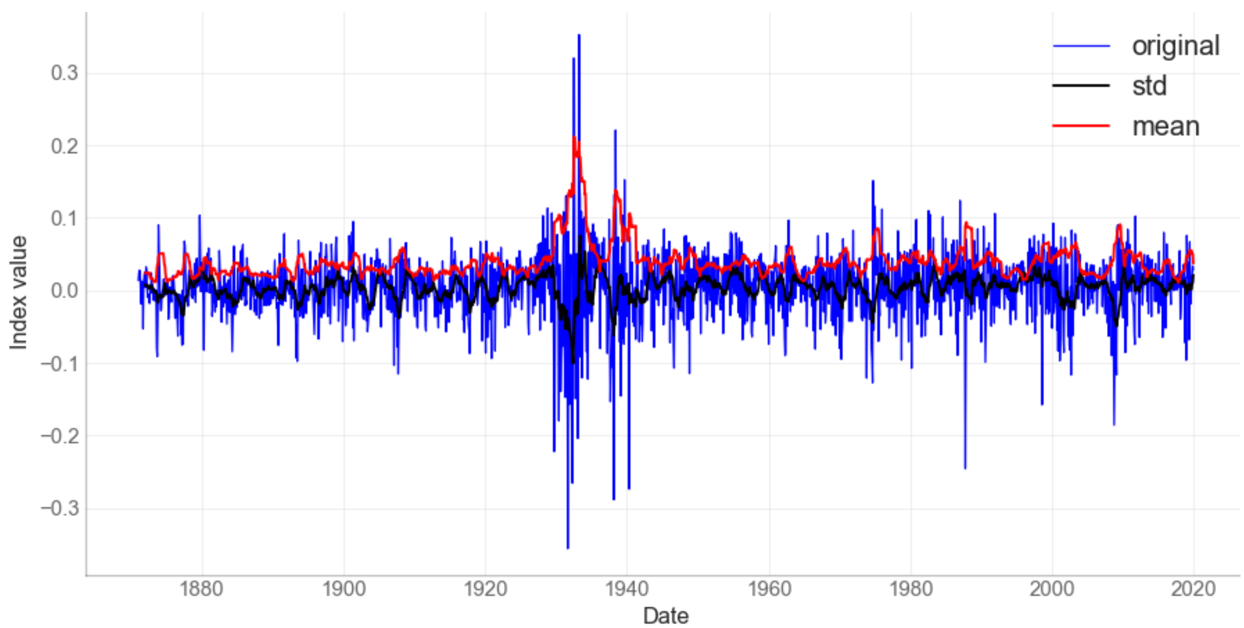



Figure 1: Log-returns stationarity

After a first visual conclusion, the group investigated the stationarity of the data by conducting two well-known tests (Augmented-Dickey-Fuller and KPSS), most commonly used in the academic literature. The Augmented Dickey-Fuller test allows for higher-order autoregressive processes by including $\Delta y_{t-p}$ in the model where $\Delta y_t$ can be written as

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \cdots + \delta_p \Delta y_{t-p} + \epsilon_t$$

The null hypothesis for the test is that the investigated data is non-stationary:

$$H_0 : \gamma = 0$$
$$H_1 : \gamma < 0$$

While the null-hypothesis of the ADF test investigates if the data is non-stationary, the null hypothesis for the KPSS test investigates if the data is stationary. As a result, one does not want to reject $H_1$. In Table below the group displayed the results for the Augmented Dickey-Fuller and KPSS test for the respective time-series (Note: a confidence level of 95% was applied while conducting the tests). By looking at the test results in Table 2, one can see that they are yielding the same outcome: both suggest that stationarity is present in the log-returns dataset.

| Test | Test Statistic | p-value | Result |
|---|---|---|---|
| Augmented Dickey-Fuller | -9.9964 | 0.000 | stationary |
| KPSS | 0.3142 | 0.463 | stationary |

Table 2: Non-Statinoarity & Stationarity Test-Results for log-returns

After confirming that the log-return time series is stationary, the group estimated an AR(1) process, whose output can be seen in Table 3 below. In this model, the value of $x$ at time $t$ is a linear function of the value of $x$ at time $t-1$. The algebraic expression of the model is as follows:

$$x_t = \delta + \phi_1 x_{t-1} + \epsilon_t$$

where:

- $\epsilon_t \overset{iid}{\sim} N(0, \sigma_w^2)$, meaning that the errors are independently distributed with a normal distribution that has mean 0 and constant variance
- Properties of the errors $\epsilon_t$ are independent of $x_t$
- An AR(1) process is (weakly) stationary, if $|\phi_1| < 1$

Below in Table 3, the group inserted the results of the fitted AR(1) process, where both intercept and regression coefficient are significant at 5%.

In conventional econometric models, the variance of the disturbance term is assumed to be constant. However many economic time series exhibit periods of unusually large volatility followed by periods of relative tranquility. In such circumstances, the assumption of a constant variance is inappropriate [1].

| | coef | std err | z | P> \|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| $\alpha_0$ | 0.0033 | 0.001 | 2.941 | 0.003 | 0.001 | 0.005 |
| $\alpha_1$ | 0.1137 | 0.024 | 4.834 | 0.000 | 0.068 | 0.160 |

Table 3: AR(1) estimation

ARCH effects are defined by Engle as such: **"a time series exhibiting conditional heteroscedasticity—or autocorrelation in the squared series—is said to have autoregressive conditional heteroscedastic (ARCH) effects"** [2]. To test the effect, we use Engle's Lagrange multiplier test.

Defining $\epsilon_t$ as the innovation of our AR(1) process, $\mathcal{F}_t$ as the filtration (information available up until time $t$) of the log return series $y_t$, its conditional variance is defined as

$$Var(y_t|\mathcal{F}_{t-1}) = Var(\epsilon_t|\mathcal{F}_{t-1}) = E(\epsilon_t^2|\mathcal{F}_{t-1}) = \sigma_t^2 \tag{2}$$

**"Thus, conditional heteroscedasticity in the variance process is equivalent to autocorrelation in the squared innovation process"** [3].

Defining the residuals as $e_t = y_t - \hat{y}_t$, Engle's Lagrange multiplier test's alternative hypothesis is the autocorrelation in the squared residuals given by a regression

$$H_1 : e_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \cdots + \alpha_m e_{t-m}^2 + u_t$$

Where $u_t$ is a white noise process. The null hypothesis is thus

$$H_0 : \ \alpha_0 = \alpha_1 = \cdots = \alpha_m = 0$$

Testing for our data, we find the results in Table 4

| | Test Statistic | p-value | Result |
|---|---|---|---|
| | 466.371 | 6.968e-83 | ARCH effect |

Table 4: Lagrange multiplier test on residuals of AR(1)

Given the p-value, the group rejects the null-hypothesis, meaning that **there is strong evidence for the presence of ARCH effect.**

# 2 Estimate an AR(1)-GARCH(1,1) model of the same series. Is the GARCH(1,1) sufficient to capture all the volatility clustering in the data?

Most of the time, it is assumed that the volatility of a time series is a constant. Obviously this is not true. Financial time series often exhibit a behaviour that is known as volatility clustering, which means that high volatility is usually followed by a high-volatility period, and this is true for low volatility, which is usually followed by a low volatility period [4]. Econometricians call this autoregressive conditional heteroskedasticity [5]. The GARCH(p,q) model generalizes ARCH(p) by considering lagged conditional variances.

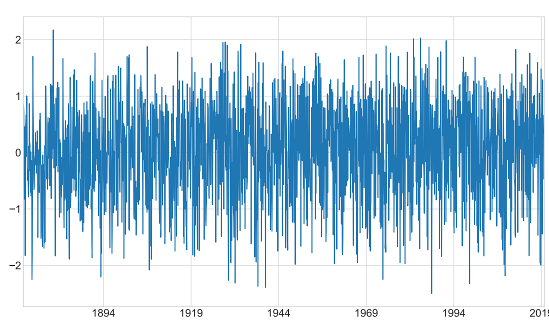
## 2.1 Methodology

To check if the fitted GARCH(1,1) captures all the volatility clustering, the group applies the following methodology, explained by Prof. Rodrigues during class:

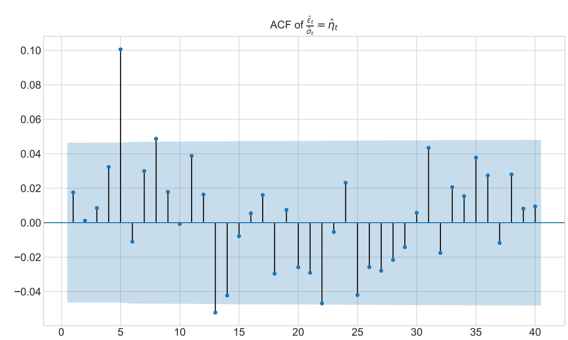
1. Compute the residuals ($\hat{\epsilon}_t$) from the fitted AR(1) process of the previous question
2. Predict the in-sample standard deviation ($\hat{\sigma}_t$) with the fitted GARCH(1,1) process
3. In the GARCH model, given than $\epsilon_t = \sigma_t \eta_t$ where $\eta_t$ is a strong white noise process ($\eta_t \sim \mathcal{N}(0, \sigma^2)$), with no correlation between its values at different times), we can easily check if the GARCH(1,1) captures all the volatility clustering: that is, checking if $\frac{\hat{\epsilon}_t}{\hat{\sigma}_t} = \hat{\eta}_t$ is indeed a white noise process.

## 2.2 Results

We observe in figures 2a and 2b that while the process **appears** to be white noise with null mean, additional features (i.e., serial correlation) are still present as shown in the ACF.




(a) In-sample estimation of standardized residuals $\hat{\eta}_t$




(b) Autocorrelogram function of standardized residuals $\hat{\eta}_t$

We then run a Ljung-Box test to ensure whether the estimated $\hat{\eta}_t$ is independently distributed or not. The Ljung-Box test examines the following hypothesis:

$$H_0: \text{ Data are independently distributed}$$
$$H_A: \text{ Data exhibit serial correlation}$$

And the results:

| $n$ | $Q_n$ | pval | $n$ | $Q_n$ | pval | $n$ | $Q_n$ | pval | $n$ | $Q_n$ | pval |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.5525 | 0.4572 | 11 | 30.0296 | 0.0015 | 21 | 43.6694 | 0.0025 | 31 | 59.1729 | 0.0016 |
| 2 | 0.5551 | 0.7576 | 12 | 30.5102 | 0.0023 | 22 | 47.6361 | 0.0012 | 32 | 59.73 43 | 0.0020 |
| 3 | 0.6843 | 0.8768 | 13 | 35.4089 | 0.0007 | 23 | 47.6894 | 0.0018 | 33 | 60.5074 | 0.0024 |
| 4 | 2.5554 | 0.6347 | 14 | 38.6271 | 0.0004 | 24 | 48.6636 | 0.0020 | 34 | 60.9403 | 0.0030 |
| 5 | 20.6725 | 0.0009 | 15 | 38.7377 | 0.0007 | 25 | 51.8597 | 0.0012 | 35 | 63.5393 | 0.0022 |
| 6 | 20.8893 | 0.0019 | 16 | 38.7908 | 0.0011 | 26 | 53.0645 | 0.0013 | 36 | 64.9134 | 0.0022 |
| 7 | 22.4990 | 0.0020 | 17 | 39.2561 | 0.0016 | 27 | 54.4751 | 0.0013 | 37 | 65.1670 | 0.0028 |
| 8 | 26.7528 | 0.0007 | 18 | 40.8370 | 0.0016 | 28 | 55.3214 | 0.0015 | 38 | 66.5981 | 0.0028 |
| 9 | 27.3255 | 0.00125 | 19 | 40.9371 | 0.0024 | 29 | 55.6857 | 0.0020 | 39 | 66.7193 | 0.0037 |
| 10 | 27.3265 | 0.0023 | 20 | 42.1445 | 0.0026 | 30 | 55.7457 | 0.0029 | 40 | 66.8819 | 0.0048 |

Table 5: Ljung-Box test on squared standardized residuals

These outputs showcase that the fitted GARCH(1,1) does not seem to capture all the volatility present in the data, given that the standardized residuals have autocorrelation (refer to the ACF and Ljung-Box).

In addition to that, the group also conducted a test to check for normality using the Jarque-Bera test:

$$H_0: \text{ The data is normally distributed}$$
$$H_A: \text{ The data is not normally distributed}$$

The test statistic is 31.97 and the corresponding p-value is 0. It means that there is sufficient evidence to say that the standardized residuals have a skewness and kurtosis significantly different from a normal distribution [6].

**In conclusion, we can infer with confidence that the GARCH(1,1) does not capture all the volatility clustering.**

# Appendix to Question 2

Since Question 2 only asks if a GARCH(1,1) process is sufficient to capture all of the volatility, the group only did a computation for this model. However, by looking at the autocorrelation function in Figure 2b, one can clearly see that there is serial correlation (additional features) in the data which is not captured by the GARCH(1,1) process.

Therefore, one ought therefore to check if the GARCH(1,1) has the optimal BIC and AIC. We observe in Table 6 that the selection criterion do not agree: GARCH(1,1) has the lowest

| GARCH | BIC | AIC |
|---|---|---|
| GARCH(p: 1, q: 1) | -6406.399336 | -6428.348030 |
| GARCH(p: 3, q: 3) | -6388.579267 | -6432.476657 |
| GARCH(p: 2, q: 2) | -6381.929125 | -6414.852167 |
| GARCH(p: 4, q: 4) | -6331.020462 | -6385.892199 |
| GARCH(p: 5, q: 5) | -6303.180275 | -6369.026359 |
| GARCH(p: 6, q: 6) | -6277.088688 | -6353.909119 |
| GARCH(p: 7, q: 7) | -6250.295466 | -6338.090246 |
| GARCH(p: 8, q: 8) | -6225.230635 | -6323.999762 |
| GARCH(p: 9, q: 9) | -6203.803398 | -6313.546872 |
| GARCH(p: 10, q: 10) | -6180.321293 | -6301.039115 |

Table 6: GARCH estimations

BIC and GARCH(3,3) has the lowest AIC.

We argue that by checking different GARCH models between (1,1) and (3,3), one could find a GARCH(p,q) that captures all the volatility clustering in the data.

# 3 Test for leverage effects (you should estimate an alternative model and perform the LR test)

Leverage effects are defined as an asymmetric reaction of volatility to returns. Chorro illustrates it in the following manner: "volatility rising more rapidly when returns are negative than positive" [7]. In other words, an asymmetric response to news is supported by theory, in which case negative shocks have a larger effect on the conditional variance. The heuristic explanation is that a negative shock raises the debt-equity ratio, thereby increasing leverage and consequently risk. It is this leverage effect that suggests bad news leads to a greater increase in conditional variance than good news.

ARCH and GARCH models assume symmetric volatility. It can be understood as positive and negative financial news having a symmetric impact on volatility [8], and that is caused by only considered the absolute size of the news (through taking its square in the ARCH/GARCH equation).
Relaxing this assumption, we consider the GJR-GARCH(1,1,1) model:

$$\sigma_t^2 = \omega + \alpha_1 \epsilon_{t-1}^2 + \gamma_1 I_{t-1} \epsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2 \tag{3}$$

Where $\sigma_t^2$ is the conditional variance at time $t$, $\omega$ the intercept, $\epsilon_t^2 = \sigma^2 z_t$ the error term composed of a time-dependent $\sigma_t^2$ and a strong white noise process $z_t$, $\gamma_1$ the scale of the asymmetric volatility, $\sigma_{t-1}^2$ the conditional variance at time $t-1$ and most importantly the indicator function $I_{t-1}$ that considers the leverage effect [9]

$$I_{t-1} = \begin{cases} 1, & \text{if } \epsilon_{t-1} < 0 \text{ negative news} \\ 0, & \text{if } \epsilon_{t-1} \geq 0 \text{ positive news} \end{cases}$$

Note that if $\gamma = 0$, the GJR-GARCH becomes a GARCH(1,1). As for the interpretation of the coefficient [9],

- $\gamma > 0$: negative shocks increase the volatility more than positive shocks
- $\gamma < 0$: positive shocks increase the volatility more than negative shocks

Refer to Table 7 for the fitted model.

| | coef | std err | z | P> \|z\| | [0.025 | 0.975] |
|---|---|---|---|---|---|---|
| omega | 0.00007 | 0.000023 | 3.092 | 1.991000e-03 | 2.547e-05 | 1.137e-04 |
| alpha[1] | 0.09640 | 0.024670 | 3.908 | 9.311000e-05 | 4.805e-02 | 0.145 |
| gamma[1] | 0.06370 | 0.036650 | 1.739 | 8.208000e-02 | -8.107e-03 | 0.136 |
| beta[1] | 0.83840 | 0.028020 | 29.925 | 9.412000e-197 | 0.783 | 0.893 |

Table 7: GJR-GARCH(1,1,1) estimation

We observe that the leverage effects are significative at 10%, giving a first indication regarding its significance. Additionally, since the coefficient is positive, it indicates that negative shocks increase more the volatility than positive shocks, which is as expected.

A better method to evaluate it, is to perform the likelihood ratio test. It is used to compare two models, where one has an additional potential confounder variable. The comparison is based on the likelihood functions [10].

The null hypothesis is that the original model fits the data better, compared to the alternative model which has an additional variable [10].

$$D = -2log(\frac{\mathcal{L}_0}{\mathcal{L}_a}) \tag{4}$$

Where $D$ is the test statistic, which follows a chisquare with degrees of freedom equal to difference in number of parameters of the two models [11]. In our case, we have 1 degrees of freedom.

Given that the GJR-GARCH model offers exactly the same as the GARCH in addition to the leverage effect, we expect it to fit the data better. What is important is whether additional feature is significative. Refer to Table 8 for our results.

| | Test Statistic | p-value | Result |
|---|---|---|---|
| | 7.1600 | 0.00745 | Alternative model fits better |

Table 8: Likelihood ratio test results

Using the likelihood ratio test, we compute a p-value indicating the significance of the additional features. We find that the additional leverage effect component of GJR-GARCH relative to GARCH is significative at 1%.

**There is strong evidence for the leverage effect in the Index log returns time series.**

# Part 2

Consider the paper from Bhargava and Malhotra: you are requested to replicate the paper between 2000 and 2020 for 4 market indexes of your choosing. You may extract the required data from Bloomberg.

## 1 Methodology

For the purpose of this replication, we use the *"Global Equity Valuations Researcher Dataset"* provided by *Siblis Research*, which contains Price-Earnings-Ratio data, stock market indexes as well as further fundamentals for several developed and developing countries. We reproduce the paper using data ranging from December 1999 to December 2020.

### 1.1 Initial Regression Specifications

We begin the replication of Bhargava and Malhotra (2006) [12] by considering the regression equations to investigate the relationship between stock prices, P/E (*Price-Earnings* ratios as well as monthly index yields. This last point is an important differentiating factor between the replication and the original paper, as we do not employ the earnings yield (which is simply the inverse of the P/E-ratio), as this could potentially introduce pesky endogeneity effects that could be difficult to control for.

$$price_t = \alpha = \beta \times \left(\frac{P}{E}\right)_t + \epsilon_t \tag{5}$$

$$yield_t = \alpha + \beta \times \left(\frac{P}{E}\right)_t + \epsilon_t \tag{6}$$

The variables are specified as follows:

1. $\underline{yield_t}$: This is the monthly index log return as calculated by:

$$yield_t = log(price_t) - log(price_{t-1})$$

2. $\underline{price_t}$: This is the index price at time $t$. Here, we can consider the stock market indexes for the USA (S&P 500), the United Kingdom (FTSE 100), the territory of Hong Kong (Hang Seng) as well as the Bovespa Index for Brazil. For this exercise, we consider only end-of-month price levels.

3. $\underline{P/E}$: This is the Price-Earnings-Ratio, which is defined as the Market Capitalization of an index divided by the aggregate earnings of the firms contained in it. As per

the replication paper, the beginning-of-month price is used. To accommodate this requirement, the series were shifted by 1, so that the end-of-month Price-Earnings-Ratio is established as the beginning-of-month Price-Earnings-Ratio of the subsequent month.

Before reporting the regression results, we will visualize the historical P/E-ratios for the four indexes selected.

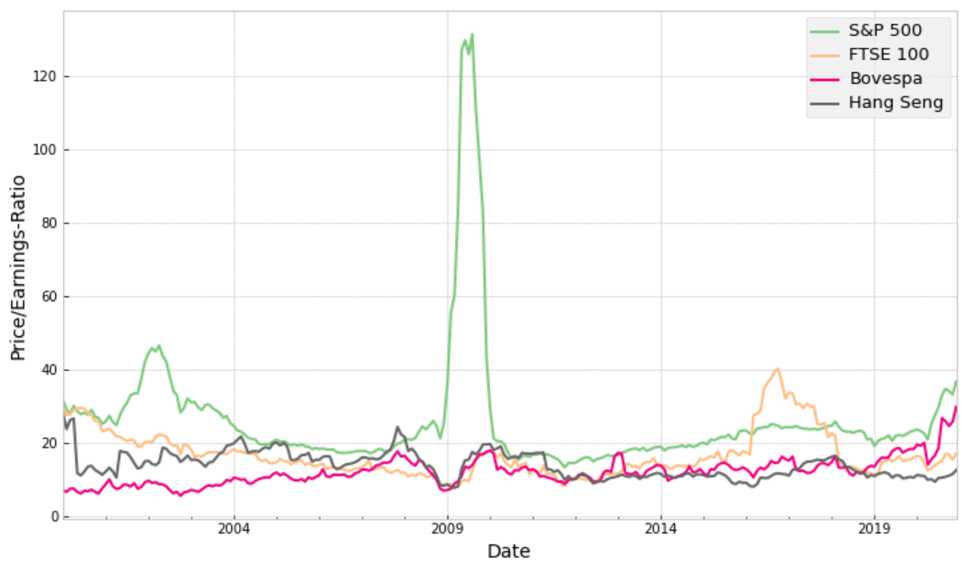



Figure 3: Historical P/E-Ratios

From figure 3 we can observe some idiosyncrasies of the time series. Very much apparent in the anomaly happening in 2009 and involving the S&P 500, during which the P/E-ratio of this index spiked to more than 120. The FTSE 100 experienced another spike around 2016, which would roughly correspond to the timeframe of the *Brexit*-referendum. What we can observe for the recent years, is a rise in the P/E-ratio for all indexes, although this rise is much less pronounced for the Hang Seng index.

## 1.2 Regression results

In tables 9 and 10 the regression results are reported. Regarding the coefficient estimates for regression 5, it appears that they are highly significant at the 1%-level. Except for the Brazilian *Bovespa* index however, the $R^2$ values are generally quite low. This means that the robustness of the results is questionable.

For regression 6, the picture is quite different. None of the coefficients in any of the four regressions is significant. $R^2$ values are near zero or negative. If there exists a relationship between monthly log returns on a market index, it was not detected within the regression framework used here.

| Index | $\alpha$ | $\beta$ | Adjusted $R^2$ |
|---|---|---|---|
| S&P 500 | 1843.8693*** | -7.2377*** | 0.030 |
| | (75.885) | (2.434) | |
| FTSE 100 | 0.1137*** | 0.024*** | 0.059 |
| | (172.328) | (9.709) | |
| Bovespa | -1.799e+04*** | 5696.56*** | 0.651 |
| | (3294.700) | (262.822) | |
| Hang Seng | 2.567e+04*** | -413.37*** | 0.066 |
| | (1371.426) | (95.683) | |

[1] Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 9: Regression results from regression 5

| Index | $\alpha$ | $\beta$ | Adjusted $R^2$ |
|---|---|---|---|
| S&P 500 | -0.0003 | 0.0002 | 0.000 |
| | (0.005) | (0.000) | |
| FTSE 100 | 0.0028 | -0.0002 | -0.003 |
| | (0.007) | (0.000) | |
| Bovespa | 0.0120 | -0.0004 | -0.004 |
| | (0.016) | (0.001) | |
| Hang Seng | 0.0147 | -0.0009 | -0.001 |
| | (0.015) | (0.001) | |

[1] Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 10: Regression results from regression 6

## 2 Empirical Analysis

Tables 11 and 12 report the results for the Durbin-Watson statistic and the White test for the residuals of regressions 5 and 6 (that is with the P/E ratio as the independent variable, and subsequent price and yield the dependent variables).

For regression 5, as the Durbin-Watson statistic values show (outside the interval 1.5-2.5), in all indexes there is enough evidence to reject the null hypothesis of no correlation among the residuals and, therefore, it can be concluded that the residuals of all indexes are autocorrelated.
Similarly, from the Lagrange Multiplier's p-values obtained when performing the White test, it can be concluded that the residuals of all indexes present heteroscedasticity (since their p-values are very close to zero and, therefore, the null hypothesis of no heteroscedasticity is

rejected). However, this is not the case for the Brazilian *Bovespa* index, which appears to have a high p-value enough to not reject the null of no heteroscedasticity in the residuals.

For regression 6, the Durbin-Watson statistic values of all indexes appear to be between 1.5 and 2.5, so there is enough evidence to not reject the null hypothesis and claim that the residuals are not autocorrelated.

Regarding the White test, this time all the Lagrange Multiplier's p-values obtained for all the indexes seem to be high enough to not reject the null hypothesis of no heteroscedasticity in the residuals, except for the S&P 500, which p-value is very low and rejects the null, concluding that its residuals present heteroscedasticity.

| Index | DW | White test (LM p-value) | Result |
|---|---|---|---|
| S&P 500 | 0.0166 | 0.0002 | AC and HET |
| FTSE 100 | 0.0589 | 0.0001 | AC and HET |
| Bovespa | 0.2718 | 1.4579 | AC and no HET |
| Hang Seng | 0.0651 | 0.0006 | AC and HET |

[1] Note: AC and HET stand for autocorrelation and heteroscedasticity, respectively

Table 11: DW and White Tests results on the residuals of regression 5

| Index | DW | White test (LM p-value) | Result |
|---|---|---|---|
| S&P 500 | 1.8528 | 0.0037 | no AC and HET |
| FTSE 100 | 1.9620 | 0.4341 | no AC and no HET |
| Bovespa | 1.7540 | 0.0984 | no AC and no HET |
| Hang Seng | 1.8219 | 0.1606 | no AC and no HET |

[1] Note: AC and HET stand for autocorrelation and heteroscedasticity, respectively

Table 12: DW and White Tests results on the residuals of regression 6

Given the autocorrelation shear found in table 11, we examine all four indices in all three time series price, yield, and P/E ratio for stationarity. For this purpose, we use the Philips-Perron unit root test. It uses the following test regression:

$$\Delta y_t = \beta D_t + \pi_1 y_{t-1} + ... + \pi_p y_{t-p} + u_t$$
$$H_0 = \pi_1 = ... = \pi_p = 0; \ existence \ of \ unit \ root$$

In using this test, we follow the methodology of the authors of the underlying paper. Apart from that, this test proves to be stronger than comparable tests like the ADF test. So it does correct for any serial correlation and heteroskedasticity in the errors $\hat{\epsilon}_t$ non-parametrically by modifying the Dickey Fuller test statistics. Phillip's and Perron's test statistics can be viewed as Dickey–Fuller statistics that have been made robust to serial correlation by using the Newey–West (1987) heteroskedasticity-and autocorrelation-consistent covariance matrix estimator. In contrast to the methodology of the authors who use four lags for the PP test without justification, our lag selection is based on the procedure that Ny and Perron [13] have presented for the optimal selection of lags.

$$MIC(k) = \ln \hat{\sigma}_k^2 + \frac{C_T(\tau_T(k) + k)}{T - k_{max}} \; with$$
$$\tau_T(k) = \frac{1}{\hat{\sigma}_k^2} \hat{\beta}_0^2 \sum_{t=k_{max}+1}^{T} y_{t-1}^2 \; and$$
$$C_T = \ln T - k_{max}$$

We use five lags for the test, which results in a good size like power of the test. To underline the importance of the selection of the lags we have shown the test for a time series in appendix B. The decisions for or against discarding $H_0$ are drastic. Table 13 shows the results of the test with the specifications, with intercept; with intercept and trend; and without intercept or trend.

| Index | Specification | S&P 500 | Bovespa | Hang Seng | FTSE 100 |
|---|---|---|---|---|---|
| Price | Intercept | 2.291 | -0.083 | -1.617 | -1.807 |
| | Int. With Trend | -0.462 | -2.097 | -3.518** | -2.851 |
| | None | 2.835 | 1.470 | 0.131 | -0.238 |
| Yield | Intercept | -14.660*** | -13.929*** | -14.701*** | -15.891*** |
| | Int. With Trend | -14.826*** | -13.900*** | -14.675*** | -15.874*** |
| | None | -14.594*** | -13.821*** | -14.713*** | -15.924*** |
| P/E | Intercept | -3.701*** | -0.869 | -4.748*** | -2.676* |
| | Int. With Trend | -3.704** | -2.574 | -5.202*** | -2.618 |
| | None | -2.015* | 0.911 | -1.759* | -1.392 |

[1] 1% critical values for intercept, intercept with trend, and no intercept or trend are: -3.46, -4.00, -2.57 respectively

[2] (*),(**),(***) refer to rejecting the unit root hypothesis at 1%, 5% and 10% significant level

Table 13: Test for Unit Root and stationarity: Phillips-Perron Test

The results of the Philip-Perron test show that for all indices and specifications for yield $H_0$ is rejected with a significance of 1%. Thus, there is no unit root and the time series are stationary. The same applies to the P/E ratio of the S&P 500 and the Hang Seng index, but with less significance. While $H_0$ for the price series of all four indices cannot be rejected under any specification and a unit root and non-stationarity exists, except for Hang Seng.
If the series are of the same order, either $I(0)$ or $I(1)$ in the case of a unit root indicates that the series are eventually cointegrated. In the presence of cointegration implies an error correction representation of the respective interacting series. To test this, we follow Johansen's error correction methodology. A theoretical proposal to study the interaction of variables that are cointegrated comes from Engle and Granger. They describe how these two variables

are in a long-time equilibrium with each other. For this purpose, a vector autoregressive model can be used, which shows the relationship between a change in this period and a change in the previous period. Furthermore, the model can be used to describe how far a system is outside its equilibrium. If no cointegration is present, a VAR model can be used. We use the Johansen test to test the linear relationship of the time series for cointegration.This allows testing for more than one cointegration.The results are shown in table 14.

| Index | Price Regressed on P/E | | Yield Regression on P/E | | Critical Value | |
|---|---|---|---|---|---|---|
| | Eigenvalue | Max-Eigen Statistic | Eigenvalue | Max-Eigen Statistic | 5% Critical Value | 1% Critical Value |
| S&P 500 | | | | | | |
| None | 0.059 | 15.25** | 0.472 | 160.39** | 14.26 | 18.52 |
| At most 1 | 0.017 | 4.39 | 0.020 | 5.18 | 3.84 | 6.63 |
| Bovespa | | | | | | |
| None | 0.140 | 37.85** | 0.581 | 218.19** | 14.26 | 18.52 |
| At most 1 | 0.030 | 7.764 | 0.087 | 22.96 | 3.84 | 6.63 |
| Hang Seng | | | | | | |
| None | 0.030 | 7.64 | 0.623 | 244.81** | 14.26 | 18.52 |
| At most 1 | 0.011 | 2.89 | 0.025 | 6.23 | 3.84 | 6.63 |
| FTSE 100 | | | | | | |
| None | 0.016 | 4.04 | 0.553 | 202.19* | 14.26 | 18.52 |
| At most 1 | 0.006 | 1.39 | 0.006 | 1.56 | 3.84 | 6.63 |

[1] (**) Max-Eigen statistics show two cointegrating equations
[2] (*) Max-Eigen statistics show one cointegrating equation
[3] The rest of the regressions reject cointegration

Table 14: Johansen Cointegration Tests

The results of the Johansen test show that two cointegrating relationships exist between price and P/E ratio as well as yield and P/E for the S&P 500 and the Bovespa. Two conitegrating relationships between Yield and P/E ratio are also significant for the Hang Seng with 5%. A cointegration equation was found for the relationship between yield and P/E ratio for the FTSE100. All other series show no cointegration.

The VECM models were estimated for the combinations of regressions of the indexes where some evidence of cointegration was found. Tables 15 and 16 show the VECM results for both the relation between price levels and P/E ratio, and between yield and P/E ratio for the S&P 500 and Bovespa indexes, respectively. Similarly, table 17 shows the VECM results only for the relation between yield and P/E ratio for both the Hang Seng and FTSE 100 indexes.

Furthermore, the Granger Causality Test was performed for all pairs of regressions, concluding the following:

1. For the S&P 500 index the P/E ratio does not Granger cause the price, but the P/E ratio Granger causes the yield.

2. For the Bovespa index, the P/E ratio Granger causes both the price and the yield.
3. And finally, for both the Hang Seng and FTSE 100 indexes, the P/E ratio Granger causes the yield.

| Independent | Dep Price | Dep P/E | Independent | Dep Yield | Dep P/E |
|---|---|---|---|---|---|
| D[Price(-1)] | 0.0099*** | -0.0254 | D[Yield(-1)] | 69.3545*** | -0.0175 |
| | (0.000) | (0.699) | | (0.000) | (0.904) |
| D[Price(-2)] | −0.0093*** | −0.1863*** | D[Yield(-2)] | 53.0289*** | -0.1148 |
| | (0.001) | (0.007) | | (0.000) | (0.413) |
| D[Price(-3)] | −0.0067** | -0.0252 | D[Yield(-3)] | 33.4660*** | -0.0166 |
| | (0.023) | (0.723) | | (0.000) | (0.897) |
| D[Price(-4)] | −0.0101*** | -0.0603 | D[Yield(-4)] | 9.5343 | 0.0288 |
| | (0.001) | (0.393) | | (0.171) | (0.781) |
| D[Price(-5)] | 0.0008 | -0.0436 | D[Yield(-5)] | 10.0285** | 0.0789 |
| | (0.797) | (0.552) | | (0.040) | (0.279) |
| D[P/E(-1)] | 0.5611*** | 0.0505 | D[P/E(-1)] | 0.4947*** | 0.0006 |
| | (0.000) | (0.973) | | (0.000) | (0.504) |
| D[P/E(-2)] | -0.0444 | -0.4574 | D[P/E(-2)] | -0.0298 | -0.0004 |
| | (0.486) | (0.765) | | (0.627) | (0.641) |
| D[P/E(-3)] | 0.5266*** | 0.3536 | D[P/E(-3)] | 0.5131*** | 0.0001 |
| | (0.000) | (0.798) | | (0.000) | (0.867) |
| D[P/E(-4)] | −0.3922*** | 0.3102 | D[P/E(-4)] | −0.3798*** | -5.339e-05 |
| | (0.000) | (0.845) | | (0.000) | (0.955) |
| D[P/E(-5)] | 0.0479 | -0.1193 | D[P/E(-5)] | 0.0645 | 0.0004 |
| | (0.426) | (0.934) | | (0.278) | (0.640) |
| Constant | 1.8280** | -29.0678 | Constant | 1.8958*** | 0.0013 |
| | (0.019) | (0.121) | | (0.000) | (0.832) |
| GC Test | $\chi^2 = 1.277$ | | GC Test | $\chi^2 = 96.603$ | |

Note: $^{*} p < 0.1$; $^{**} p < 0.05$; $^{***} p < 0.01$

Table 15: VECM Results for the S&P 500 index

| Independent | Dep Price | Dep P/E | Independent | Dep Yield | Dep P/E |
|---|---|---|---|---|---|
| D[Price(-1)] | 0.0002*** | 0.1531** | D[Yield(-1)] | 3.5385 | 0.1791 |
| | (0.000) | (0.022) | | (0.135) | (0.306) |
| D[Price(-2)] | -1.002e-06 | $-0.1818$** | D[Yield(-2)] | 3.9676* | 0.0836 |
| | (0.961) | (0.035) | | (0.063) | (0.597) |
| D[Price(-3)] | -1.586e-06 | $-0.1430$* | D[Yield(-3)] | 3.5803* | 0.0359 |
| | (0.938) | (0.096) | | (0.051) | (0.791) |
| D[Price(-4)] | 1.905e-05 | $-0.2619$*** | D[Yield(-4)] | 3.5672** | -0.0295 |
| | (0.351) | (0.002) | | (0.018) | (0.791) |
| D[Price(-5)] | $-6.77e-05$*** | -0.0788 | D[Yield(-5)] | 1.1557 | -0.0682 |
| | (0.001) | (0.365) | | (0.309) | (0.416) |
| D[P/E(-1)] | 0.0915 | 166.1823 | D[P/E(-1)] | 0.0114 | 0.0035 |
| | (0.177) | (0.559) | | (0.865) | (0.482) |
| D[P/E(-2)] | 0.0152 | 410.5300 | D[P/E(-2)] | -0.0444 | 0.0056 |
| | (0.817) | (0.135) | | (0.497) | (0.247) |
| D[P/E(-3)] | 0.0475 | 805.5815*** | D[P/E(-3)] | 0.0127 | 0.0080 |
| | (0.470) | (0.003) | | (0.846) | (0.100) |
| D[P/E(-4)] | 0.0954 | 65.5806 | D[P/E(-4)] | 0.0233 | 0.0006 |
| | (0.156) | (0.816) | | (0.726) | (0.905) |
| D[P/E(-5)] | 0.0207 | 201.5764 | D[P/E(-5)] | 0.0293 | 0.0003 |
| | (0.705) | (0.379) | | (0.692) | (0.951) |
| Constant | 0.5848** | -175.1562 | Constant | -0.0036 | 0.0264 |
| | (0.043) | (0.885) | | (0.989) | (0.167) |
| GC Test | $\chi^2 = 15.078$ | | GC Test | $\chi^2 = 166.307$ | |

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

Table 16: VECM Results for the Bovespa index

For the remaining combinations of regression for the indexes, where no cointegrating relationship could be found, VAR models were estimated using the *Akaike Information Criterion* (AIC) as the metric, based on which the optimal lag length was selected. Results are reported in table 18 below. For the Hang Seng Index, we find significant coefficients for the trend and the first lag of the price when *Price* is the dependent variable. When $PE$ is the dependent variable, however, the constant, the first and second lag of *Price* and the first lag of $PE$ are significant. For the FTSE 100, the two significant coefficients are the first lag of *Price* and the third lag of $PE$, when the Price-Earnings Ratio is the dependent variable.

The Granger Causality Test (using the Wald-Test) resulted in the confirmation that neither for the Hang Seng Index, neither for the FTSE 100 Index the Price-Earnings Ratio Granger-causes the Index price level. The relevant statistics remain well below the necessary thresholds for that hypothesis.

| Independent | Dep Yield | Dep P/E | Independent | Dep Yield | Dep P/E |
|---|---|---|---|---|---|
| D[Yield(-1)] | -3.7718 | -0.1145 | D[Yield(-1)] | 11.4104** | -0.1154 |
| | (0.132) | (0.526) | | (0.011) | (0.446) |
| D[Yield(-2)] | -2.7466 | -0.0772 | D[Yield(-2)] | 10.2187** | -0.1428 |
| | (0.211) | (0.625) | | (0.014) | (0.309) |
| D[Yield(-3)] | $-3.4722^{*}$ | 0.0170 | D[Yield(-3)] | 7.7054** | -0.1485 |
| | (0.072) | (0.902) | | (0.038) | (0.236) |
| D[Yield(-4)] | $-2.8512^{*}$ | 0.0493 | D[Yield(-4)] | 2.7650 | -0.0929 |
| | (0.076) | (0.670) | | (0.366) | (0.368) |
| D[Yield(-5)] | -1.1593 | 0.0119 | D[Yield(-5)] | 1.6190 | -0.0311 |
| | (0.300) | (0.882) | | (0.454) | (0.669) |
| D[P/E(-1)] | 0.0212 | 0.0019 | D[P/E(-1)] | 0.0292 | -0.0002 |
| | (0.732) | (0.678) | | (0.640) | (0.912) |
| D[P/E(-2)] | 0.0213 | -0.0049 | D[P/E(-2)] | 0.0231 | 0.0014 |
| | (0.632) | (0.124) | | (0.711) | (0.515) |
| D[P/E(-3)] | 0.0066 | 1.568e-05 | D[P/E(-3)] | 0.3604*** | 0.0024 |
| | (0.882) | (0.996) | | (0.000) | (0.216) |
| D[P/E(-4)] | -0.0154 | 0.0043 | D[P/E(-4)] | -0.0034 | -0.0010 |
| | (0.725) | (0.170) | | (0.957) | (0.629) |
| D[P/E(-5)] | -0.0172 | 0.0043 | D[P/E(-5)] | 0.0402 | -0.0006 |
| | (0.693) | (0.171) | | (0.523) | (0.773) |
| Constant | 0.9326*** | 0.0291 | Constant | 0.6532*** | 0.0024 |
| | (0.000) | (0.102) | | (0.004) | (0.748) |
| GC Test | $\chi^2 = 239.758$ | | GC Test | $\chi^2 = 77.114$ | |

Note: $^{*}\, p < 0.1$; $^{**}\, p < 0.05$; $^{***}\, p < 0.01$

Table 17: VECM Results for the Hang Seng (left) and FTSE 100 (right) indexes with Yield and P/E ratios

| Independent | Dep Price | Dep P/E | Independent | Dep Price | Dep P/E |
|---|---|---|---|---|---|
| Price(-1) | 0.9632*** | 0.0006*** | D[Price(-1)] | -0.0297 | 0.0027*** |
| | (0.0652) | (0.0000) | | (0.0639) | (0.0004) |
| Price(-2) | 0.1108 | $-0.0005^{***}$ | D[Price(-2)] | -0.0018 | -0.0002 |
| | (0.0991) | (0.00005) | | (0.0721) | (0.0004) |
| Price(-3) | -0.0983 | -0.00007 | D[Price(-3)] | -0.0117 | -0.0006 |
| | (0.1057) | (0.00008) | | (0.0722) | (0.0004) |
| Price(-4) | -0.0983 | 0.00004 | D[Price(-4)] | - | - |
| | (0.1052) | (0.00008) | | - | - |
| Price(-5) | -0.0241 | -0.00006 | D[Price(-5)] | - | - |
| | (0.0801) | (0.00006) | | - | - |
| P/E(-1) | -70.8742 | 0.9223*** | D[P/E(-1)] | -5.600972 | 0.0346 |
| | (67.2129) | (0.0484) | | (11.0846) | (0.0606) |
| P/E(-2) | -21.4710 | -0.0404 | D[P/E(-2)] | 5.8245 | 0.0420 |
| | (88.7595) | (0.0640) | | (11.0877) | (0.0607) |
| P/E(-3) | 91.1165 | 0.0108 | D[P/E(-3)] | 15.8402 | 0.2811*** |
| | (87.4983) | (0.0631) | | (9.9341) | (0.0543) |
| P/E(-4) | 60.8967 | 0.0367 | D[P/E(-4)] | - | - |
| | (81.0571) | (0.0584) | | - | - |
| P/E(-5) | -53.0842 | -0.0372 | D[P/E(-5)] | - | - |
| | (52.6176) | (0.0379) | | - | - |
| Constant | 1090.587 | 1.9129*** | Constant | - | - |
| | (562.3493) | (0.4056) | - | - | |
| Trend | 7.775** | -0.0028 | Trend | - | - |
| | (2.7686) | (0.0019) | - | - | - |
| GC Test | $\chi^2_{(5)} = 5.114$ | | GC Test | $\chi^2_{(3)} = 3.085$ | |

[1] Note: $^{*}\,p < 0.1$; $^{**}\,p < 0.05$; $^{***}\,p < 0.01$

Table 18: VAR Results for the Hang Seng index (left) and FTSE 100 (right)

# 3 General Conclusion

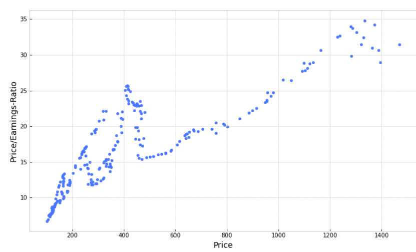
The analysis of the relationship between yields, price-earnings-ratios as well as index prices, as conducted above, has brought forward a mixed picture regarding the replication of the paper by Bhargava and Malhotra (2006). The main factor here was the difference in the distributions of the variables considered. While in Bhargava and Malhotra (2006, 88), the PE-ratios used showed a clear upward trend, this has not been the case for the sample considered here. In fact, the sample from 2000 to 2020 has shown several anomalies for the stock indexes considered, which, by visual inspection, seem to be associated with the 2016 Brexit referendum for the FTSE 100 and with the aftermath of the Financial Crisis in case of the S&P 500 for the US.

According to the results from the previous VECM models, Granger causality between the P/E ratio and the yield for all four indexes can be claimed, whereas there seems to be no Granger causality between the P/E ratio and the price for any other index, except for the Brazilian *Bovespa* index. Concerning the VAR results, no model could validate the directionality of Granger causality from *Price-Earnings-Ratios* to *Price*.

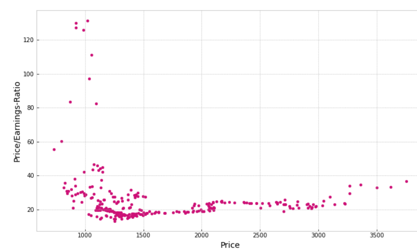
To make the reported results more robust, we considered several enhancements to the methodology, such as the lag order selection for the Phillips-Perron tests as proposed in [13] as well as VAR lag order selection by means of the minimization of the Akaike Information Criterion (AIC).

# A Scatterplots (Price vs. Price-Earnings-Ratio)

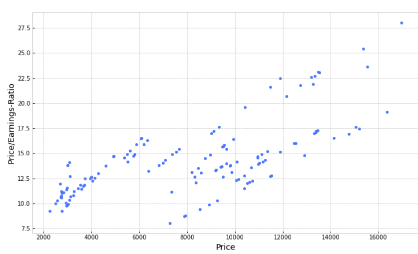
The following three pairs of scatterplots show *Price* on the x-Axis and $Price - Earnings - Ratio$ on the y-Axis. Note that the shape of all scatterplots differs markedly between the time ranges considered. Whereas the earlier shapes in the 1980s and 1990s shows a clear upward trend, this is not the case for the subsequent period that is considered in this exercise.



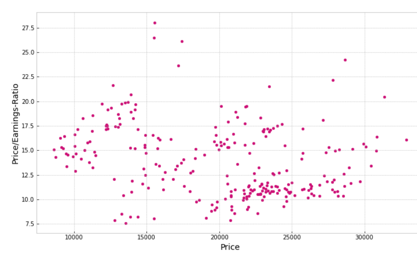

(a) S&P 500 (1980 - 1999)



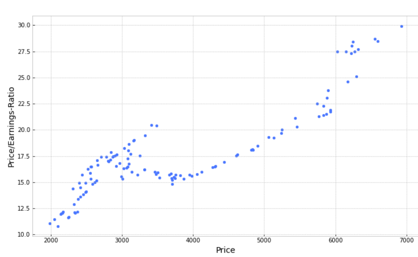

(b) S&P 500 (2000-2020)



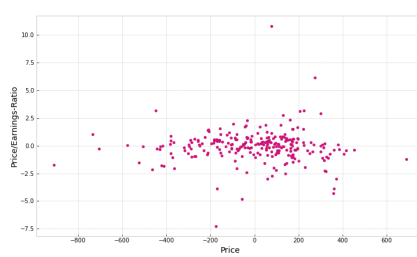

(a) Hang Seng (1988 - 1999)




(b) Hang Seng (2000-2020)




(a) FTSE 100 (1990 - 1999)




(b) FTSE 100 (2000-2020)

# B Lag choice Phillips-Perron Test

| S&P500 Price | | |
|---|---|---|
| Lags | Test Statistic | Critical value 1%, 5%, 10% |
| 1 | -1.468 | -2.57, -1.94, -1.62 |
| 2 | -1.689* | -2.57, -1.94, -1.62 |
| 3 | -1.864* | -2.57, -1.94, -1.62 |
| 4 | -1.964** | -2.57, -1.94, -1.62 |
| 5 | -2.015** | -2.57, -1.94, -1.62 |
| 6 | -2.032** | -2.57, -1.94, -1.62 |
| 7 | -2.006** | -2.57, -1.94, -1.62 |
| 8 | -1.903* | -2.57, -1.94, -1.62 |

[1] *** refers to rejecting the null hypothesis on a 1% significance level
[2] ** refers to rejecting the null hypothesis on a 5% significance level
[3] * refers to rejecting the null hypothesis on a 10% significance level

Table 19: Phillips-Perron Test for Price of S&P500 with different lags with neither trend nor intercept

# References


[1] Michael Jansson. "Problem Set 1 Solutions and Pierre Yared". In: *Massachusetts Institute of Technology* (2004).

[2] Robert F Engle. "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation". In: *Econometrica: Journal of the econometric society* (1982), pp. 987–1007.

[3] Mathworks. *Engle's ARCH Test*. https://www.mathworks.com/help/econ/engles-arch-test.html?w.mathworks.com. Accessed on 2021-03-02.

[4] Yuxing Yan. *Python for Finance: Financial modeling and quantitative analysis explained*. Vol. 2nd edition. Packt Publishing, Apr. 2014.

[5] Christoph Hanck et al. *Introduction to Econometrics with R*. University of Duisburg-Essen, Sept. 2020.

[6] *"How to Perform a Jarque-Bera Test in Python"*. https://www.statology.org/jarque-bera-test-python/. Accessed on 2021-02-15. July 2020.

[7] Chorro Christophe et al. "Testing for Leverage Effect in Financial Returns". In: *SSRN Electronic Journal* (Jan. 2014). DOI: 10.2139/ssrn.2590269.

[8] N/A. *ARCH/GARCH Volatility Forecasting*. https://goldinlocks.github.io/ARCH_GARCH-Volatility-Forecasting/. Accessed on 2021-03-03.

[9] Christoffer Dahlvid and Per Granberg. *The Leverage Effect - Uncovering the true nature of U.S. asymmetric volatility*. eng. Student Paper. 2017.

[10] Joanna Diong. *Likelihood ratio test: relevance and application*. https://scientificallysound.org/2017/08/24/the-likelihood-ratio-test-relevance-and-application/. Accessed on 2021-03-03. 2017.

[11] statsmodels. *Python stats model documentation for LR test*. https://www.statsmodels.org/stable/generated/statsmodels.regression.linear_model.OLSResults.compare_lr_test.html. Accessed on 2021-03-03. 2020.

[12] Vivek Bhargava and D. K. Malhotra. "Do Price-Earnings Ratios Drive Stock Values?" In: *The Journal of Portfolio Management* (2006), pp. 86–92. DOI: 10.3905/jpm.2006.661379.

[13] Serena Ng and Pierre Perron. "Lag Length Selection and the Construction of Unit Root Tests with Good Size and Power". In: *Econometrica* (2001), pp. 1519–1554.