# Time Series Analysis and Forecasting using ARMA Models: A Case Study of the Philippine Stock Exchange

**Brynx Junil Alegarbes, Benjamin Louis Ang, Jeremy Marcus Tan**

## Abstract

The accurate prediction of stock market movements and the forecasting of future trends on the Philippine Stock Exchange (PSE) present significant challenges for investors, traders, and financial institutions. This paper aims to address the issue by exploring the performances of time series models in forecasting the Philippine Stock Exchange Composite Index in the next 3 years using historical monthly data of about 36 years from February 1987 to April 2022. Time series from the original data was found to be non-stationary in the Augmented Dickey-Fuller Test. After log transformation and differencing, the series became stationary. Also, by Ljung-Box Test, the series is serially correlated. Time series models to be tested were found to be AR(1), MA(1), and ARMA(1,1), after evaluating the lags in the ACF and PACF correlograms. The best time series model was determined to be AR(1) after having the lowest AIC value $(-870.63)$. In forecasting, the AR(1) model was found to have a MAPE value of $6.62\%$, indicating highly accurate forecasts, and residuals were found to have a zero mean and non-correlation. Thus, the paper suggests that stakeholders and investors could use AR(1) to form their expectations and decisions.

**Keywords:** *ARMA, time series, forecasting, stocks, Philippine Stock Exchange*

## 1 Introduction

Time series analysis is a way of analyzing a sequence of data points collected over an interval of time. It is used to uncover patterns, understand dynamics, and make predictions in various economic and financial domains. Among its models is the Autoregressive Moving Average (ARMA) model, which is adept at handling stationary time series. In this study, the researchers fit an ARMA model to the Philippine Stock Exchange data, with the aim of creating a model which uses historical price data to predict future market movements.

The PSE, or Philippine Stock Exchange, is the national stock exchange in the Philippines. It operates as a self-regulatory body, facilitating as a market where securities can be bought and sold in a fair, efficient, transparent, and orderly manner. It is a crucial component of the country's financial system, providing a dynamic platform for trading a diverse range of securities. Economists consider the stock market as a predictor for the overall economy of a country since movements in the stock market incorporate forward-looking information as stock prices reflect the expectations for future earnings [3]. Hence, forecasting future stock market performance is of great interest to investors, economists, and policymakers.

A stock market index is a hypothetical portfolio of investments that is meant to represent a segment of the stock market and serves as a benchmark for the entire market. In the United States, the S&P 500 index and the NASDAQ composite index are two of the most popular market indices. In the Philippines, the PSEi is the main index for the Philippines Stock Exchange. It tracks 30 of the largest blue-chip companies on the Philippines Stock Exchange like Jollibee Foods Corp (JFC) and Ayala Corporation (AC). The index is market capitalization weighted which means that it gives more weight to larger companies as compared to smaller ones. The PSEi provides a good representation of the stock market performance of the largest companies in the Philippines. The primary objective of this research is to conduct an analysis of the time series behavior exhibited by the Philippine Stock Exchange Composite Index (PSEi).

This study serves as a fundamental stepping stone in comprehending the intricate behavior of the Philippine Stock Exchange data. It provides a launching pad for future research, enabling the exploration of more advanced modeling approaches that encompass additional factors and account for the high volatility exhibited by the PSE. We strive to empower stakeholders with quantitative insights, enhance investment decision-making, and contribute to a deeper understanding of the dynamic relationship between the stock market and the broader economy.

### 1.1 Statement of the Problem

The Philippine Stock Exchange (PSE) plays a crucial role in the country's financial landscape, serving as a platform for investment and capital formation. However, accurately predicting stock market movements and forecasting future trends on the PSE remains a challenging task. The lack of robust time series modeling and forecasting techniques tailored specifically for the Philippine stock market hinders investors, traders, and financial institutions from making well-informed decisions. Therefore, the study aims to:

1. Gather and analyze historical data from the Philippine Stock Exchange;
2. Develop and implement time series models (AR, MA, and ARMA) for predicting stock exchange index and forecasting market trends on the Philippine Stock Exchange;
3. Assess the performance of the different time series models and identify the best model to capture stock market dynamics and generate reliable forecasts; and
4. Contribute to the existing body of knowledge in the field of financial forecasting offering insights and empirical evidence specific to the Philippine Stock Exchange and its unique market dynamics.

### 1.2 Scope and Limitations

This study only models data for the Philippine Stock Exchange Index. Thus, the model's forecasts are only applicable to 30 of the largest companies in the Philippines. The predictions cannot be generalized to other companies not in the index, which are mostly mid-cap and small-cap stocks. Moreover, the study only used monthly closing prices in training the model. The weekly and daily closing prices may have contained additional information which would have improved the predictive power of the model.

It is crucial to acknowledge the unique challenge posed by the high volatility of the PSE. The presence of substantial price fluctuations and rapid market movements necessitates the need for sophisticated modeling techniques to capture the intricate dynamics within the data. While the ARMA model provides a solid foundation, it is important to recognize the ARMA model's assumption of constant variance, which makes its fit for stock market data not ideal [1].

Moreover, the stock market is influenced by a multitude of factors, including macroeconomic indicators, investor sentiment, geopolitical events, and global market conditions. The interplay of these complex dynamics presents additional challenges in accurately modeling the behavior of the Philippine Stock Exchange. Other studies used different time series models like Autoregressive Conditional Heteroskedasticity (ARCH) to analyze volatility and the GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model to forecast volatile and nonlinear price data [6].

## 2 Methods

This study uses the autoregressive moving average model (ARMA) in forecasting the future prices of the PSEi based on discrete time series data.

The ARMA model is a type of regressive model with two components: autoregression and moving average. The autoregression model predicts future values based on past values. It relies on the assumption that past values are indicative of future values. On the other hand, the moving average model states that future values can be predicted using current and past error terms. These error terms are assumed to be white noises, independent from each other, having a mean of zero.

### 2.1 Mathematical Theorems, Definitions, and Tests

**Definition 1.1.** A **time series**, denoted by $\{X_t : t \in T\}$ or $\{X_t\}$, is a set of observations, each one being recorded at a specific time.

**Definition 1.2.** A **time series model** for the observed data $\{x_t\}$ is a specification of the joint distributions of a sequence of random variables $\{X_t\}$ of which $\{x_t\}$ is postulated to be a realization.

**Definition 1.3.** $\{X_t\}$ is an **ARMA$(p, q)$** process if $\{X_t\}$ is stationary and if for every $t$,

$$X_t - \phi_1 X_{t-1} - \cdots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \cdots + \theta_q Z_{t-q},$$

where $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$ and the polynomials $(1 - \phi_1 z - \cdots - \phi_p z^p)$ and $(1 + \theta_1 z + \cdots + \theta_q z^q)$ have no common factors.

**Definition 1.4.** Let $x_1, \ldots, x_T$ be observations of a time series.

- The **sample ACVF** is $\hat{\gamma}(h) = \frac{1}{T} \sum_{t=h+1}^{T} (x_t - \bar{x})(x_{t-h} - \bar{x}), 0 \leq h < T - 1$.
- The **sample ACF** is $\hat{\rho}(h) = \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)} = \frac{\sum_{t=h+1}^{T} (x_t - \bar{x})(x_{t-h} - \bar{x})}{\sum_{t=1}^{T} (x_t - \bar{x})^2}, 0 \leq h < T - 1$.

**Definition 1.5** Let AR($p$) be an autoregressive process of order $p$ and MA($q$) be a moving average process of order $q$. Then,

1. PACF cuts off at lag $p$
2. ACF cuts off at lag $q$

**Ljung-Box (1978): Test for Autocorrelation**

1. $H_0 : \rho_1 = \rho_2 = \cdots = \rho_m = 0$
   $H_1 : \rho_i \neq 0$ for some $i \in \{1, \ldots, m\}$where $m \approx lnT$
2. $Q(m) = T(T + 2) \sum_{h=1}^{m} \frac{\hat{\rho}_h^2}{T-h}$
3. Reject $H_0$ if $Q(m) > \chi^2(m, 1 - \alpha)$

**Augmented Dickey-Fuller Test (1979): Test for Stationarity**
Consider the characteristic equation of an AR process.
$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$

1. $H_0 : \phi_i = 1$ for some $i \in \{1, \ldots, p\}$
   $H_1 : \phi_i \neq 1$ for all $i \in \{1, \ldots, p\}$
2. Test Statistic: $\hat{\tau}_\mu = \hat{\phi}_1^* / \hat{SE}(\hat{\phi}_1^*)$
3. Given $\alpha = 0.05$. If $\hat{\tau}_\mu < -2.86$, reject $H_0$. We conclude that the process is stationary.

### 2.2 Dataset

This study uses publicly available data from Yahoo Finance for the historical monthly closing prices of the PSEi. The website has complete monthly data from February 1987 to April 2022, so this is chosen as the dataset for this study. The CSV file from Yahoo Finance had 432 rows, one for each month, and 5 columns: Date, Open, High, Low, Close, and Adjusted Close. The CSV file was uploaded to RStudio, where the closing prices were extracted and made into a time series for further analysis.

### 2.3 Model Assumptions

The time series data has to satisfy several assumptions before being modeled. The following is a list of assumptions for the ARMA model:

1. The time series must be weakly stationary.
2. The time series must exhibit autocorrelation.

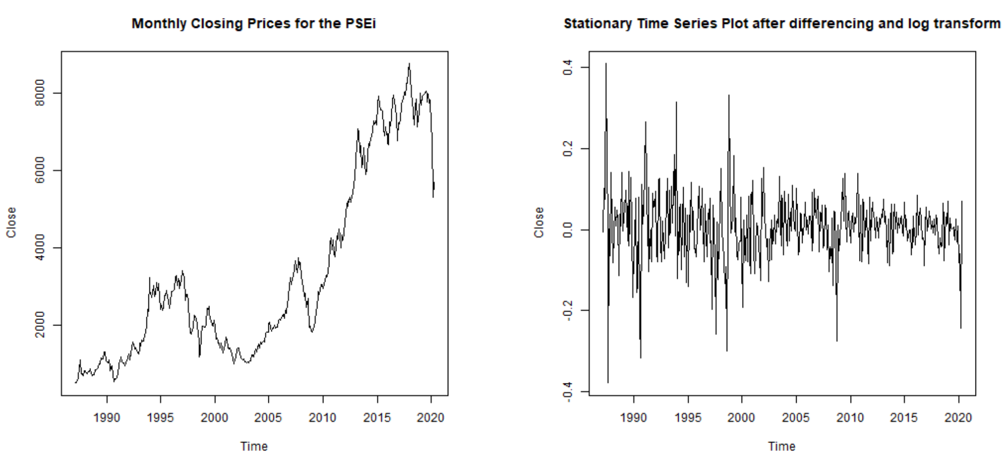
### 2.4 Pre-Processing

The researchers used R programming language throughout the analysis. In specific, the following libraries in RStudio are used: TSA, tseries, lmtest, tidyverse, forecast, quantmod, ggplot2, and xts.

The dataset was split into training and testing sets in order to evaluate the forecasts produced. At the discretion of the researchers, the training data was from February 1987 to April 2020 while the testing data was from May 2020 until April 2022.

Two pre-processing steps were done on the training data to make it stationary: log transformation and differencing.

- A logarithmic transformation was done to the data. This transformation is often used to stabilize the variance of the time series and reduce the impact of extreme values. It can also help in linearizing relationships and making patterns in the data more apparent.
- After the log transformation, differencing on the data was performed. Differencing involves computing the differences between consecutive observations in the time series. This step is commonly used to remove trends or seasonal patterns from the data, making it stationary.




(a) Original Data

(b) Pre-Processed Data

Figure 1: Time Series Plot

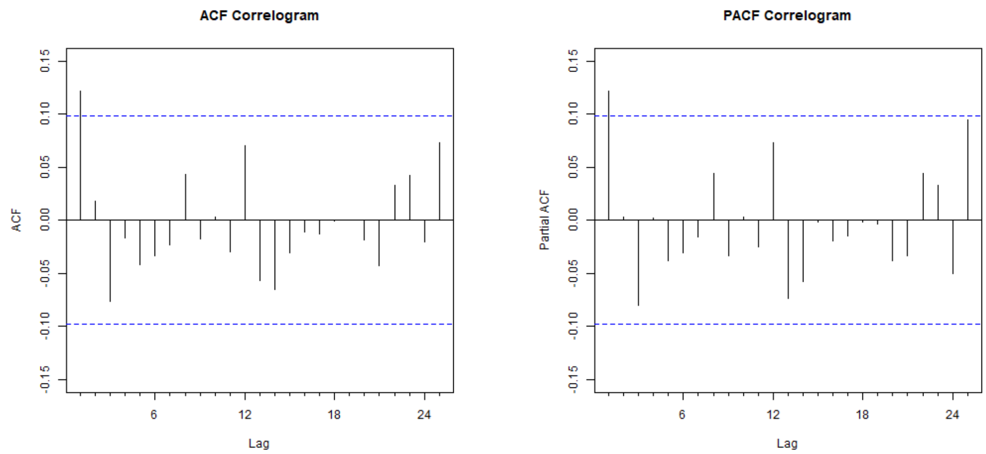
Notice that after pre-processing, the time series data now looks stationary. The stationarity must still be verified by objective tests in the next subsection.

### 2.5 Check for Stationarity and Autocorrelation

The preprocessed data was tested using the Augmented Dickey-Fuller (adf) and Ljung-Box Test. The $p$-value given by the adf and the Ljung-Box test were $0.01$ and $0.01453$, respectively. Hence, the pre-processed data was stationary and serially correlated.

### 2.6 Identifying Lag Orders and Best Model

Following that, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the pre-processed data were analyzed. The ACF and PACF exceeded the $95\%$ confidence threshold at lag 1 for both metrics as shown in Figure 2. In both correlograms, the ACF and PACF converged to zero after lag 1. Thus, the candidate models to be tested are AR(1) and MA(1). The ARMA(1,1) model was also tested.




(a) ACF (b) PACF

Figure 2: Correlograms for pre-processed data.

The three candidate models were then compared using the Akaike Information Criterion (AIC) as shown in Table 1.

Table 1: AIC Values for Candidate Models

| Model | AIC |
|---|---|
| AR(1) | $-870.63$ |
| MA(1) | $-870.41$ |
| ARMA(1,1) | $-868.63$ |

The model with the lowest AIC was found to be AR(1), which has an AIC of $-870.63$, in comparison to MA(1) with an AIC of $-870.41$ and to ARMA(1,1), which had an AIC of $-868.63$.

## 3 Results and Discussion

### 3.1 Best Model

The best model for the log-transformed and differenced PSEi data is the AR(1) model with the equation as follows:

$$X_t = 0.0061 + 0.1213X_{t-1} + Z_t,$$

where $Z_t \sim$ W,N$(0, \sigma^2)$.

### 3.2 Checking Residuals

The researchers then analyzed the residuals $\{Z_t\}$ of the best-fitting model. The distribution of the residuals was graphed using RStudio. Looking at the distribution, the residuals roughly have a mean of zero. The residuals were tested using the Ljung-Box test, which produced a $p$-value of $0.921$. Thus, the residuals are not autocorrelated.

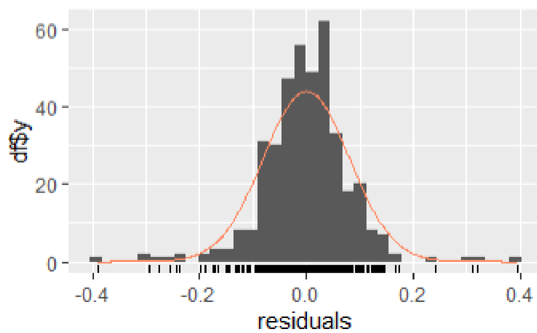
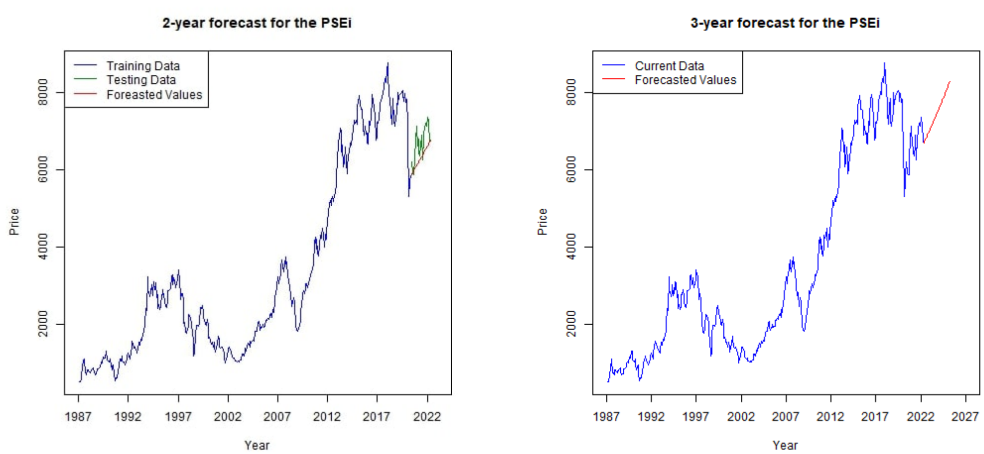



Figure 3: Distribution of the AR(1) model.

### 3.3 Forecasting and Evaluation against the Test Data

Using the AR(1) model, the researchers forecasted the closing price of the PSEi over two years, from May 2020 to April 2022, the same time period as the testing data. The operations (differencing and log transform) were inversed to get back to the original data. The forecasted values are in line with the actual test values for the 2-year period. The mean absolute percent error (MAPE) between the actual and forecasted values was 6.62%. A MAPE value less than 10% indicates a highly accurate forecast [4].




(a) Comparison from 2020 to 2022

(b) Forecasted values from 2022 to 2025

Figure 4: Forecasts for the PSEi

Finally, the AR(1) model was used to forecast the prices of the PSEi from May 2022 to April 2025. The graph indicates that the PSEi is predicted to steadily increase over the following years.

## References


[1] A. T. Andreea-Cristina Petrica, Stelian Stancu. Limitation of arima models in financial and monetary economics. *Theoretical and Applied Economics*, 23(4(609)):19–42, 2016.

[2] P. Brockwell and R. Davis. *Introduction to Time Series and Forecasting*. Springer, 3rd edition, 2016.

[3] B. Comincioli. The stock market as a leading indicator: An application of granger causality. *University Avenue Undergraduate Journal of Economics*, 1, 1996.

[4] C. Kasemset, N. Sae-haew, and A. Sopadang. Multiple regression model for forecasting quantity of supply of off-season longan. *Chiang Mai University Journal of Natural Sciences*, 13:391–402, 12 2014.

[5] Yahoo Finance. Psei index. `https://rb.gy/tu5ti`, May 2023.

[6] Y. Yu. A study of stock market predictability based on financial time series models. *Mobile Information Systems*, 2022.


## A Appendix

See attached R Notebook, PSEI_Monthly.rmd for codes and the corresponding dataset, PSEI.PS.csv.