Time Series Analysis Techniques for Financial Time Series: An Overview

Maria Pia Bellini

January 2023

1 Abstract

Financial time series are commonly volatile and non-stationary, making it challenging to forecast and manage risk using many traditional techniques. As a result, financial time series analysis is crucial in various fields, including risk management and the development of effective management strategies. This essay provides a brief overview of various models and techniques that have been developed over the years, which do not rely on particular assumptions regarding the distribution of data, such as stationarity and the absence of noise. These techniques range from more traditional approaches like bootstrapping to more recent models like Neural Networks.

2 Introduction

Financial time series are sets of observations collected over a period of time, usually over regular intervals of time. Those observations reveal financial measures such as the return of risk, the beta with the market, the price of a stock and so on.

2.1 Why does Financial Time Series Analysis matter?

Developing useful techniques to analyze financial time series is a problem of interest to many subjects, such as:

- **Investors**: individual investors, institutional investors, other financial istitutions are interested in the analysis of financial time series with to make investment decisions, manage their portfolios, maximizing the returns while minimizing risks;
- **Traders**: all kinds of traders, (e.g. day, swing, position traders...) rely on financial time series analysis to identify short-term trading opportunities, execute trades, and generate profits from market movements;

- **Risk Managers**: risk managers within financial institutions, asset management firms, and investment funds analyze financial time series to assess and mitigate various types of risks, including market risk, credit risk, liquidity risk, and operational risk;

- **Economists**: Economists and economic researchers use financial time series analysis to study the relationships between financial variables, macroeconomic indicators, and economic policies. They seek to understand how financial markets impact the broader economy and vice versa;

- **Regulators**: financial regulators and government agencies analyze financial time series to monitor market trends, detect anomalies or irregularities, and enforce regulatory compliance to maintain the stability and integrity of financial markets;

- **Academics**: researchers and academics in finance, economics, and related fields use financial time series analysis for empirical research, testing financial theories, and publishing academic papers that contribute to the understanding of financial markets and their behavior.

The mentioned subjects are just some of the several figures interested in the analysis of this kind of measure. Financial measures are a matter of interest to a multitude of fields, and understanding and trying to predict their behavior is crucial for several reasons.

- **Investment Decision Making**: helping investors make informed decisions about buying, selling, or holding financial assets such as stocks, bonds, or commodities. By analyzing historical price movements and patterns, investors can identify potential investment opportunities and assess risk;

- **Risk Management**: financial institutions and investors can manage risk effectively. By studying past price movements, volatility, and correlations among different assets, risk managers can develop strategies to hedge against potential losses and stabilize investment portfolios;

- **Forecasting**: forecast future price movements and market trends. By applying statistical models and technical analysis techniques to historical data, analysts attempt to predict future price movements, which is essential for making strategic decisions and formulating investment strategies;

- **Market research**: providing valuable insights into market dynamics, investor sentiment, and macroeconomic factors influencing asset prices. This information is crucial for market research, helping analysts understand market trends, identify emerging opportunities, and anticipate changes in market conditions;

- **Economic Analysis**: financial time series analysis is also used in economic research to study the relationships between financial variables and macroeconomic factors. By examining historical data on indicators such as GDP, inflation, interest rates, and employment, economists can analyze the impact of economic policies and external events on financial markets and vice versa;

2.2 Why Financial Time Series Analysis are so difficult to analyze?

The crucial factor that makes financial time series tedious to analyze is the intrinsic characteristics of the financial metrics, such as volatility clustering is one of them. Volatility is a measure of the degree of variation of a variable over time, for example, high volatility of a security implies high fluctuations in its returns. Since this measure is not directly measurable, statistical theory and methods play an important role. [13]
This is the time series of the price of the JPMorgan stock from November 2004 to September 2023, reported as an example:

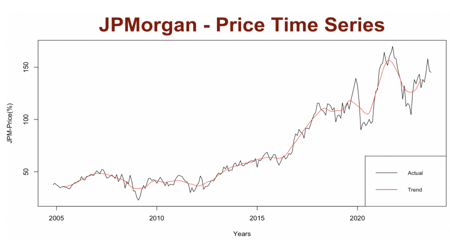

Figure 1: Time Series of JPMorgan Price from Nov 2004 to Sept 2023

Financial time series volatility tends to cluster, which implies the volatility behavior in a certain period of time will be similar in the following period, rather than being randomly distributed over time.
Stationarity implies that the statistical properties of a time series, such as mean, variance, and autocovariance, remain constant over time. However, financial time series often exhibit non-stationary behavior, which can pose challenges for the application of traditional time series analysis techniques.
The presence of volatility clustering has several implications for the application of traditional time series analysis techniques:
Volatility clustering leads to heteroscedasticity, meaning that the variance of the time series is not constant over time. Traditional time series analysis techniques, such as linear regression or autoregressive models, assume constant variance (homoscedasticity) for their underlying assumptions to hold. In the presence of volatility clustering, these techniques may produce biased estimates and unreliable forecasts.
In addition, this property often results in autocorrelation in the time series, where the values of the series at different time points are correlated and non-normality in the distribution of financial returns or prices. This violates the

assumption of independence and normal distribution of the data required for many traditional time series analysis techniques, such as the ARIMA (AutoRegressive Integrated Moving Average) model. Autocorrelation can lead to incorrect parameter estimates and unreliable model diagnostics.
In general traditional models that do not account for volatility clustering, meaning that they may fail to capture important features of the data, leading to poor forecasting performance and unreliable risk management.
Leptokurtosis is a statistical term used to describe the shape of a distribution, particularly in the context of financial returns. It indicates a departure from the normal distribution, resulting in a distribution with fatter tails compared to the normal distribution.
In simpler terms, leptokurtosis refers to the tendency for financial returns to exhibit more extreme values, both positive and negative, than would be expected under a normal distribution.
The **leverage effect**, also known as asymmetry, describes a phenomenon observed in financial markets where the relationship between stock returns and changes in volatility is not symmetrical. It suggests that stock returns tend to have a negative correlation with changes in volatility. This means that when bad news or negative events is affecting the market, volatility are expected to increase. Conversely, when there is good news or positive events, volatility tends to decrease.
Overall, the violation of the stationarity hypothesis in financial time series necessitates the adoption of specialized modeling techniques that can accommodate non-stationarity and capture the underlying dynamics of the data more effectively than traditional time series analysis techniques based on stationary assumptions.
Surely we could consider several approaches and models, but in this essay, we will focus on three of them: ARCH/GARCH models, non-linear regression, spectral analysis, and deep learning hybrids.

3 Related work

3.1 ARCH/GARCH models and the EGARCH

AL-Najjar D. (2016) [1], analyzes the returns and prices of Amman Stock Exchange general index data (AISEI) from January 1st, 2005 to December 31st, 2014.
We will understand why even though ARCH-GARCH models are suitable for financial time series, they may fail to capture leptokurtic distributions and leverage effect, so we will use an implemented model like the EGARCH.
ARCH (Autoregressive Conditional Heteroscedasticity) and GARCH (Generalized Autoregressive Conditional Heteroscedasticity) models are widely used in financial time series analysis, particularly in the context of stock markets, to model and forecast the volatility of asset returns. These models are essential

for handling certain characteristics of financial time series, such as volatility clustering and heteroscedasticity as mentioned before.
The **ARCH model** assumes that the conditional variance of a time series is a function of past squared errors or residuals from a mean model. The model is typically denoted as ARCH(p), where 'p' represents the number of lags of squared errors included in the model.
The general equation of the model is:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + ... + \alpha_p \varepsilon_{t-p}^2$$

where:

- σ_t^2 is the conditional variance of the time series at time 't';
- ε_t represents the error term at time 't';
- $\alpha_0, \alpha_1, ..., \alpha_p$ are the parameters to be estimated;

The parameters of the ARCH model are typically estimated using maximum likelihood estimation (MLE) or other optimization techniques. The model can be estimated iteratively by first estimating the mean model (e.g., ARMA) and then estimating the ARCH parameters based on the residuals of the mean model.
The ARCH model acknowledges that the level of volatility in a financial time series can vary over time, and this variation is influenced by past errors or deviations from the mean. By incorporating information about past errors into the model, the ARCH model can capture the changing nature of volatility in the data.
Some limitations of the ARCH model that led to the implementation of the GARCH model, though, include the assumption of constant parameters. The limited persistence in volatility dynamics, since it only accounts for the impact of past squared errors on current volatility leads to inefficiency in capturing these long memory effects.
The **GARCH model** allows a longer memory and a more flexible lag structure. It extends the ARCH model by incorporating lagged conditional variances in addition to squared errors. The model is denoted as GARCH(p, q), where 'p' represents the number of lags of squared errors and 'q' represents the number of lags of conditional variances included in the model.
The general equation of the model can be written as:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

where:

- $\alpha_0, \alpha_1, ..., \alpha_p, \beta_1, ..., \beta_q$ are the parameters to be estimated;

Once the parameters of the ARCH or GARCH model are estimated, the conditional variance can be forecasted for future time periods.

Many studies, such as Hsieh (1989) [7], Taylor (1994) [12] concluded that the best model to describe the data and measure the volatility is GARCH.
The problem arising from the application of these models is that both ARCH and GARCH models assume the symmetric distribution of all the shock effects on volatility. The empirical results show that this is not always the case for many stock markets in the world, depending on their different characteristics.
Many researchers provided extended forms of the ARCH model, such as the GARCH-M, provided by Engle et al. (1987), [4], with the ability to capture the feedback effect between volatility and the mean of the time series, and the GJR GARCH, Glosten et al. (1993) [5], based on the dependence of the volatility asymmetric response on the positive and negative shocks.
The implemented model we're interested in this essay though is the EGARCH (**Exponential GARCH**), Nelson (1991), [10], for which positive and negative shocks affect the variance of returns differently.
The general equation of the model can be written as:

$$ln(\sigma_t^2) = \omega + \sum_{j=1}^{p} \beta_j ln(\sigma_{t-j}^2) + \sum_{i=1}^{q} \alpha_i \left\{ |\frac{\varepsilon_{t-i}}{\sigma_{t-i}}| - \sqrt{\frac{2}{\pi}} \right\} - \gamma_i \frac{\varepsilon_{t-i}}{\sigma_{t-i}}$$

where:

- no restrictions on parameters ω, β, γ are required;

Moreover, to ensure the stationarity assumption still holds, we must ensure that β positive and less than 1, and $\gamma\}$ (an indicator of leverage effect) is negative and significant.
This model leads to more fitness accuracy in the estimation of volatility in comparison to the other ARCH models, due to its ability to show the greater impact of volatility by large shocks, and to test both symmetric and asymmetric distributions.

3.2 Non-linear Regression

Al-shaiby (2020) [2] demonstrated non-linear models are effective in predicting the stock prices within the coming period and hence determining an effective strategy to buy and sell. In particular, he analyzed the closing prices of Apple Inc. (AAPL) from October 1st, 2019 to July 8th, 2019.
Predicting stock prices and returns may be very challenging as the determination of a stock price depends on a multitude of factors, as observed for Apple Inc.:

- **The economic process**, demand, and supply fluctuations;
- The value of the organization, corresponding to its **capitalization**, that's equal to the stock price increased by the number of shares outstanding;
- **Corporate indices**, such as price-to-earnings (obtained by multiplying the price-earnings ratio by the rate of profit growth) and investment index (the value of the company that the investor will buy its shares);

Multiple linear regression models the relationship between a dependent variable and two or more independent variables. It extends the concept of simple linear regression, which models the relationship between a dependent variable and a single independent variable. This relationship is expressed as a linear equation of the form:

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n + \varepsilon$$

where:

- Y is the dependent variable;
- $X_1, ..., X_n$ are the indipendent variables;
- $\beta_0, \beta_1, ..., \beta_n$ are the parameters to estimate;
- ϵ is the error term;

The parameters $\beta_0, \beta_1, ..., \beta_n$ are estimated through the OLS method, which chooses the values of the parameters that minimize the Sum of Squared Errors (SSE). Predicting closing pricing by implementing a linear regression analysis between the stock prices and the various factors we just considered is not effective as the relationships between those variables may be highly complex.

In a **nonlinear regression model**, the observational data is modeled by a function given by a nonlinear combination of the model parameters. Such parameters are estimated with different techniques, such as Maximum Likelihood Estimator (MLE), maximizing the likelihood function, which measures the probability to observe the given data under the assumed model, and Nonlinear Least Squares (NLLS), minimizing the sum of squared differences between the observed and predicted values of the dependent variable using an iterative optimization algorithm.

A nonlinear model is hence more suitable in predicting financial variables, a suitable model can be achieved by making some initial assumptions on the parameters and iteratively adjusting them according to some model evaluation criteria (e.g. ANOVA test).

3.3 Spectral Analysis

Harmonic analysis can be effective in predicting the behavior of stock market values due to its ability to capture periodic patterns and cycles in the data. Financial markets exhibit various periodic behaviors such as daily, weekly, monthly, or yearly cycles, which can influence asset prices, trading volumes, and other financial metrics.

Escuanuela R. et al. stated that stock prices don't follow a Random Walk Movement (RWM), but follow regular cyclical movements that harmonic analysis can find [13]. Others performed spectral analysis on financial time series and proved their efficiency, such as Granger (1963), who concluded that short-term frequencies may follow a RWM, but long-term ones don't, and the latter are more relevant for the model prediction [6]. Harmonic analysis is applied

to time-stationary series, so nonstationary series are generally made stationary removing trend and seasonality, with techniques such as differencing or least squares.
The Autocorrelation function is used to analyze the behaviour of the series, the ACF is the correlation between one time series x_t and a lagged one x_{t+k}. The equation of the ACF at lag k is:

$$\rho_k = \frac{cov(x_{t,t-k})}{\sqrt{Var(x_t)Var(x_{t-k})}}$$

The object of the **spectral analysis** is to extract hidden periodicities of the series, a stationary time series can be expressed as a trigonometric function:

$$x_t = \sum_{j=1}^{n/2} A_j cos(2\pi \mathrm{f}_j + \varphi_j)$$

where:

- A_j is a coefficient;
- φ_j) is a phase factor;
- $\mathrm{f}_j = \frac{j}{n}$ are the harmonic frequencies, lying between 0 and 0.5 [11].

The goal is to identify the most influential harmonic frequencies in signal modeling, often done using **the periodogram**. While the periodogram is unbiased, it's not consistent, achieving consistency involves averaging over neighboring frequencies, which sacrifices some detail in exchange for reliability. Hence the application of RWM is not valid, while the harmonic approach validates the identification of major movements despite their high dispersion and irregularity. While it may not precisely estimate every recorded movement, it effectively captures the most significant ones.

3.4 Deep Learning Hybrids

Various empirical studies proved that applying deep learning approaches permits to achieve better prediction accuracy.
Deep learning approaches can implement nonlinear, highly complex functions that receive inputs and return outputs. The most popular deep learning techniques used in financial time series prediction are:

- **Multi-layer Perceptron (MLP)**: a simple multilayer perceptron (MLP) model typically comprises three main components: an input layer, a hidden layer, and an output layer. The input layer receives historical financial time series data as input features, representing observations at previous time steps such as historical prices, trading volumes, and economic indicators. Each input feature corresponds to a neuron in the input layer;

- **Convolutional Neural Network (CNN)**: typically consists of input layers, convolutional layers for detecting local patterns, pooling layers for dimensionality reduction, and fully connected layers for making predictions. During training, CNNs learn to identify relevant patterns and relationships within the data using supervised learning techniques. Once trained, CNNs can make predictions on future financial time series data, providing valuable insights for tasks such as forecasting stock prices and predicting market trends.;
- **Long Short-term Memory (LSTM)**: these networks capture long-term dependencies in sequential financial data like historical prices or trading volumes. They consist of input layers, LSTM layers for temporal pattern learning, and fully connected layers for predictions. LSTM networks are trained to model complex relationships, enabling accurate predictions on future financial data, and making them valuable for tasks like stock price forecasting and market trend prediction.;

A hybrid forecasting model, known as "Hybrid," blends two or more separate forecasting methods to make predictions better and fix any issues with individual methods. Many studies have shown that these hybrid models work better than using just one method, especially when predicting things over time.
Researchers have found that combining different methods, like using math or machine learning, helps make predictions more accurate. By mixing these methods, hybrid models can understand more about the data and make better guesses about what will happen in the future. [9]
Mohan B.H. et Krishna (2019) [9] reviewed the majority of deep learning hybrids for financial time series prediction from 1999 to 2019. They counted 34 different kinds of hybrids, these are the types they found:

- **MLP-based hybrids**: a proposed MLP-based hybrid was the GA+MLP+EFK: the GA (i.e. Generic Algorithm) found the optimal structure of MLP, and the localized version of the EKF (Extended Kalman Filter) is used for training. This hybrid yields better multi-step-ahead predictions but shows small sensitivity in strong fluctuations periods [3];
- **LSTM-based hybrids**: the GARCH-LSTM is a hybrid formulated to forecast stock price volatility, it combines LSTM with various GARCH models. The hybrid model effectively combined two strengths: it was able to learn sequential patterns well, and it improved the accuracy of predicting changes in stock market volatility [8];
- **CNN-based hybrids**: CNN-TA is an algorithmic trading model, it converts financial time series into 2-D images, later labeled as "Buy", "Sell" or "Hold" depending on the trend of the series. This hybrid is particularly efficient for stocks and ETFs (Exchange-Traded Funds).

Looking into detail at all the 34 kinds of hybrid deep learning models proposed in the last 20 years, it's observed the majority of them performed better than stand-alone models [9].

3.5 Which is the better technique to analyze Financial Time Series?

So, is there an optimal technique to analyze Financial Time Series that is always effective? Each method we encountered in this essay offers unique insights and capabilities, catering to different aspects of financial analysis.
The following tables summarize the characteristics of the approaches we reviewed in this essay:

Technique	Pros
ARCH	Captures the changing nature of volatility
GARCH	It allows a longer memory and a more flexible lag structure
EGARCH	Test both symmetric and asymmetric distributions
Nonlinear Models	Capture non-linear relationships between variables
Spectral Analysis	Captures periodic patterns and cycles in the data
MLP-Based Hybrids	Better multi-step-ahead predictions
LSTM-Based Hybrids	High Accuracy in predicting changes in stock market volatility
CNN-Based Hybrids	Particularly efficient for stocks and ETFs

Technique	Cons
ARCH	Inefficiency in capturing long memory effects
GARCH	Assumption of the symmetric distribution of all the shock effects on volatility
EGARCH	
Nonlinear Models	
Spectral Analysis	Does not precisely estimate every recorded movement
MLP-Based Hybrids	Small sensitivity in strong fluctuations periods
LSTM-Based Hybrids	
CNN-Based Hybrids	

In conclusion, each method offers distinct advantages in analyzing financial data. ARCH/GARCH models and non-linear regression provide insights into volatility and complex relationships, while spectral analysis uncovers periodic patterns. Deep learning hybrids, on the other hand, harness the power of advanced algorithms to deliver superior predictive accuracy. Understanding the strengths and limitations of each approach is crucial for applying the technique more suitable for each case.

4 Conclusions

Financial Time Series are difficult to analyze with traditional techniques due to their chaotic characteristics.
Empirical studies demonstrated that ARCH models are suitable for capture the changing nature of volatility in the data but may fail to incorporate long memory effects only accounting for the impact of past squared errors on current volatility. GARCH models are implemented by incorporating lagged conditional variances in addition to squared errors, that allows for longer memory and more flexible lag structure. Leptokurtosis and Leverage Effect of returns distribution may still cause problems to GARCH models, because of the assumption of the symmetric distribution of all the shock effect on volatility.
Many implementations of these models exists, but in particular we saw that the EGARCH model effectively captures leptokurtosis and leverage effect.
Multiple linear regression models are not suitable for predicting stock prices and returns, as linear functions fails to capture the relationships between them and the variety of regressors they depend on.
The application of nonlinear models and the application of model evaluation criteria such as ANOVA test had proven to be more effective in representing those relationships and providing more accurate forecasts.
RWM is not be the most suitable approach for stock market values, as it assumes stock prices follow a random motion. Spectral analysis claims the series follows regular cyclical movements, effectively estimating the most important movements, even though it does not permit us to capture every single movement.
Deep learning models are proven to be particularly effective for financial time series, as with their complex structure they are able to capture their random patterns.
Hybrid forecasting models are better performing in prediction accuracy of financial metrics, since they overcome the limits of the sand-alone models. Several hybrids have been proposed in the last 20 years, in particular the most popular are the one MLP, CNN and LSTM based.

References

[1] Dana Al Najjar. Modelling and estimation of volatility using arch/garch models in jordan's stock market. *Asian Journal of Finance Accounting*, 8:152, 05 2016.

[2] Nazar M. Al-shaiby. Using the nonlinear regression to predict the prices of stocks and its impact on the investor decisions in the financial market, 05 2020.

[3] Georgopoulos E. F. Andreou, A. S. and S. D. Likothanassis. Exchange-rates forecasting: A hybrid algorithm based on genetically optimized adaptive neural networks. *Computational Economics, 20(3), pp. 191-210.*, 2002.

[4] R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica, 50(4), 987–1007*, 1982.

[5] Jagannathan R. Runkle D. E. Glosten, L. R. On the relation between expected value and the volatility of the nominal excess return on stocks. *ournal of Finance, 48(5), 1779–1801*, 1993.

[6] C. W. Granger. The typical spectral shape of an economic variable. *Econometrica: Journal of the Econometric Society, 150-161*, 1966.

[7] D.A. Hsieh. Modeling heteroscedasticity in daily foreign-exchange rates. *Journal of Business Economic Statistics, 7, 307-317*, 1989.

[8] Won C. H. Kim, H. Y. Forecasting the volatility of stock price index: A hybrid model integrating lstm with multiple garch-type models. *Expert Systems with Applications, 103, pp. 25-37.*, 2018.

[9] Krishna Mohan B.H. A review of two decades of deep learning hybrids for financial time series prediction. *International Journal on Emerging Technologies*, 10:324–331, 01 2019.

[10] D. B. Nelson. Conditional heteroscedasticity in asset returns: A new approach. *Econometrica, 59(2), 347–370*, 1991.

[11] Ignacio Escanuela Romana and Clara Escanuela Nieves. A spectral approach to stock market performance, 2023.

[12] S. Taylor. Modelling stochastic volatility: A review and comparative study. *Mathematical Finance, 4, 183-204.*, 1994.

[13] Ruey S Tsay. *Analysis of financial time series*. John wiley & sons, 2005.