Sefirot
Financial Research

Arima Models In Finance: Predicting Markets With Historical Series

Authors:
Mallia Corrado
Croci Mattia

September, 2025

Index

1 Introduction

In the world of finance, predicting the evolution of prices, rates, and market indices is a fundamental objective for investors, analysts, and institutions. Even though markets are influenced by complex and often unpredictable factors, statistical models allow us to extract useful information from historical data. Among these tools, ARIMA models, acronym for **Autoregressive Integrated Moving Average**, represent a consolidated reference point in time series analysis. Developed by Box and Jenkins in the 1970s, they are appreciated for their clear structure, ease of interpretation, and ability to describe even complex temporal dynamics, provided that the series is made stationary.
Although they were not specifically designed for financial markets, ARIMA models are often applied to data such as log returns, where some of their assumptions are more plausible. Their main strength is simplicity combined with a solid theoretical foundation. The most evident limitation is the inability to model typical financial phenomena, such as time-varying volatility or nonlinear relationships. This article explains how ARIMA models work, how to build them step by step, and in which financial contexts they can be effectively useful, without hiding their weaknesses.

2 What are Historical Series in Finance

In finance, a historical series is a sequence of observations ordered over time related to an economic or financial variable. Classic examples include daily stock prices, exchange rates between currencies, returns of a stock index like the S&P 500, short-term interest rates, or trading volumes. Each point of the series represents the value assumed by that variable at a specific moment, and the entire sequence reflects the dynamic evolution of the phenomenon over time. The analysis of these series underpins many financial activities, from risk management to trading strategy development, asset valuation, and macroeconomic planning.
Financial historical series have some distinctive characteristics that complicate their analysis. First, they often show a trend, i.e., a long-term tendency to grow or decline, due to structural factors such as inflation, economic growth, or technological innovation. Second, although less evident than in other fields such as retail or energy, in some cases they may exhibit seasonality or cyclicality, for example linked to fiscal deadlines, investors' seasonal behaviors, or calendar effects. An even more relevant feature is high volatility, which tends to cluster over time: periods of great instability alternate with phases of relative calm, a phenomenon known as **volatility clustering**. Finally, many financial series are non-stationary, meaning that their mean, variance, or autocorrelation change over time.
Stationarity is a fundamental condition for applying ARIMA models. A stationary process is one whose statistical properties, particularly mean, variance, and autocovariances, do not depend on time but only on the distance between observations. This assumption is crucial because ARIMA models are based on the idea that the future behavior of the series can be coherently described from the relationships observed in the past. If the series is non-stationary, these relationships become unstable, and parameter estimates become unreliable, leading to misleading forecasts. For this reason, before applying an ARIMA model, it is necessary to transform the original series into a stationary series, typically through one or more differencing operations, which constitute the "I" (Integrated) component of the model. Without this preliminary step, the model would not be able to correctly capture the underlying dynamics of financial data.

3 The ARIMA Model: Structure and Components

The ARIMA model, acronym for **Autoregressive Integrated Moving Average**, is one of the pillars of time series analysis, widely used in economics, finance, and many other applied fields. Its strength lies in the ability to represent the temporal dynamics of a variable observed over time in a synthetic and structured way, such as stock prices, exchange rates, or stock index returns. What makes ARIMA particularly versatile is the fact that it does not impose a rigid form on the data but adapts to different types of temporal behavior thanks to three integer parameters: **p**, **d**, and **q**. These numbers completely define the model's structure and correspond to its three fundamental components: autoregression, integration, and moving average. The standard notation is therefore **ARIMA(p, d, q)**, and the appropriate choice of these values is the crucial step to build an effective and reliable model.

The first component, autoregression (AR), is related to parameter **p** and is based on an intuitive idea: the value that a series assumes today is not entirely independent of what happened in the past. In finance, this temporal dependence often manifests as a form of inertia in returns or prices. For example, if a stock has posted positive returns in recent days, there may be some probability that this trend continues, at least in the short term, due to mechanisms like momentum or collective investor behavior.

Parameter **p** indicates how many past values are used to explain the current value. An AR(1) model considers only the immediately preceding observation, while an AR(3) includes the last three. Increasing **p** allows capturing more complex dependence structures but also exposes to overfitting risk, especially when data are noisy, as often happens in financial markets.

The second component, called integration (I), is controlled by parameter **d** and addresses one of the most important challenges in financial series analysis: lack of stationarity. A stationary series is one whose mean, variance, and autocorrelation structure do not change over time. Stock prices, however, tend to grow in the long run, showing an upward trend that violates this condition. Volatility may also vary drastically from one period to another, for example during financial crises, making variance unstable. To apply an ARIMA model, it is necessary to transform the original series into a stationary series. This is achieved through differencing, which consists of replacing each value with the difference from the previous period. If the series has a linear trend, a single difference (**d = 1**) is often sufficient. In more complex cases, **d = 2** may be required, but in finance, especially when working with logarithmic returns (which are already a form of difference of log prices), **d** rarely exceeds 1. Parameter **d** is therefore the bridge that connects raw data to the world of stationary models, where ARIMA can operate correctly.

The third and final component is the moving average (MA), associated with parameter **q**. Despite its name, this part of the model has nothing to do with moving averages used in technical analysis. Instead, it refers to past forecast errors, i.e., unexpected shocks not captured by the model. In a financial context, an unexpected event such as a sudden monetary policy change, a sovereign default, or a surprising corporate announcement can generate a price movement that was not predictable with the information available until the previous day. These shocks, although random, can have a residual effect: the market may continue to digest the news in the following days, generating some persistence in the error. The MA component assumes that the current value of the series also depends on these past errors. Parameter **q** indicates how many of these residuals are included in the model: an MA(1) considers only the shock from the previous period, an MA(2) includes the one from two periods ago, and so on. This component is particularly useful for modeling short-term but high-impact phenomena, very common in markets.

Combined, the ARIMA(p, d, q) model merges in a single coherent structure the memory of past levels (**p**), the correction for non-stationarity (**d**), and the echo of random shocks (**q**). The result is a flexible, interpretable, and mathematically solid tool, capable of adapting to a wide range of temporal behaviors. However, its effectiveness critically depends on the correct identification of parameters, which requires preliminary analysis (such as autocorrelation plots), statistical tests (e.g., for stationarity), and a post-estimation diagnostic phase. Additionally, ARIMA assumes linearity, homoscedasticity, and normality of errors, assumptions often violated in real financial data. Despite these limits, ARIMA remains an essential reference: not so much as a definitive solution, but as a solid foundation for understanding the temporal structure of data and building more advanced models, such as hybrid ones with GARCH or machine learning techniques.