AR, MA, and ARMA Models: Historical Development, Theory and Application to Financial Time Series

Andrea Buffoli

June 30, 2025

1 Introduction

The analysis of time series models has a long and rich history, motivated by the need to understand and forecast sequential data in economics, engineering, and natural sciences. The **Autoregressive (AR) model** was introduced by Yule in 1927 (1) while studying sunspot numbers. Later, Slutsky proposed the **Moving Average (MA) model** in 1937 (2), highlighting the effects of random shocks in economics. The unification of these ideas, the **Autoregressive Moving Average (ARMA) model**, was systematically developed by Box and Jenkins in the 1970s (3), leading to the renowned Box-Jenkins methodology for time series analysis.

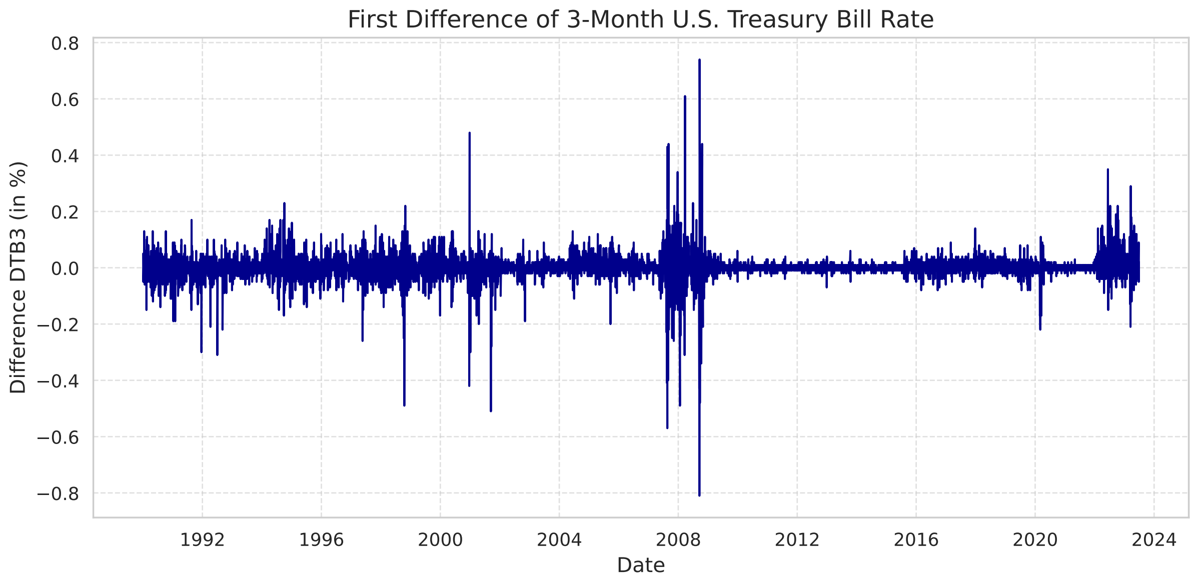

Figure 1: First difference of the 3-month U.S. Treasury Bill rate. The plot displays the daily changes in the DTB3 series from 1990 to 2025.

Today, AR, MA, and ARMA models are essential tools for the modeling and forecasting of financial time series such as stock returns, interest rates, and exchange rates.

Their ability to capture autocorrelation, mean reversion, and volatility dynamics makes them central to quantitative finance, econometrics, and risk management.

2 Mathematical Formulation

2.1 Autoregressive, Moving Average, and ARMA Models

In time series analysis, three fundamental models are widely used to represent and predict data that evolve over time: the autoregressive (AR), moving average (MA), and the combined autoregressive moving average (ARMA) models. We denote the observed time series by $\{X_t\}_{t\in\mathbb{Z}}$, where t indicates the time period.

Autoregressive Model AR(p) An autoregressive model of order p expresses the value at time t, X_t, as a linear combination of its p previous values plus a constant (intercept) and a random shock:

$$X_t = c + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t, \tag{1}$$

where c is the intercept (drift), $\phi_1, \ldots, \phi_p$ are parameters, and ϵ_t is white noise: a sequence of independent random variables with mean zero and variance σ^2. Intuitively, this model states that the present depends on the recent past, a baseline level c, and a random unpredictable component.

Moving Average Model MA(q) A moving average model of order q expresses X_t as a linear combination of a constant, the current random shock, and the past q random shocks:

$$X_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}, \tag{2}$$

where μ is the mean level, $\theta_1, \ldots, \theta_q$ are parameters, and ϵ_t is white noise. Here, the value at time t depends on the average level μ and random shocks in the current and previous q time periods.

Autoregressive Moving Average Model ARMA(p, q) The ARMA(p, q) model combines both effects:

$$X_t = c + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}. \tag{3}$$

This allows the process to capture both long-term dependencies (AR) and short-term shocks (MA).

Example: Suppose daily temperature can be explained by averaging the temperatures of the previous two days (AR(2)), plus a baseline level and a random effect for unexpected weather. Or, daily stock returns may have a non-zero average return, plus memory of previous market shocks (MA model).

2.2 Stationarity: Definition and Interpretation

Before analyzing the properties of these models, it is crucial to ensure that the process is **stationary**. **Stationarity** means that the main statistical properties of the process, such as the mean, variance, and autocorrelation, do not change over time. In other words, the process "looks the same" at any point in time: it does not exhibit trends, periodicities, or growing variances.

Formally, for an AR(p) process, stationarity is guaranteed if the roots of the characteristic polynomial

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p,$$

lie **outside the unit circle** in the complex plane; that is, for each root z^*, we have $|z^*| > 1$.

Intuitive explanation:

- The coefficients $\phi_1, \ldots, \phi_p$ determine how much past values influence the present. If their combined effect is "too strong," the process can become explosive or non-stationary: small shocks in the past can be amplified indefinitely.
- The characteristic polynomial $\phi(z)$ encodes this combined influence in a compact mathematical form.

What does "outside the unit circle" mean and why is it important?

- The *unit circle* in the complex plane is the set of complex numbers z such that $|z| = 1$.
- If any root z^* of $\phi(z)$ satisfies $|z^*| \leq 1$, the process will either "explode" (i.e., values can grow without bound) or display persistent, non-decaying oscillations.
- Only when all roots satisfy $|z^*| > 1$ do past shocks fade away over time, ensuring that the influence of any shock eventually dies out and the series remains stable and predictable in its statistical behavior.

Practical implications:

- Stationarity is required for the mathematical properties (mean, variance, autocorrelation) to be well defined and constant over time.
- Most statistical inference techniques, as well as forecasting and model selection procedures, assume the process is stationary.
- In practice, before fitting an AR(p) model to real-world data, it is common to test or enforce stationarity, for example by transforming the data (e.g., differencing to remove trends).

Example: Consider an AR(1) model:

$$X_t = c + \phi_1 X_{t-1} + \epsilon_t.$$

Here, the characteristic polynomial is $\phi(z) = 1 - \phi_1 z$. The root is $z^* = 1/\phi_1$. For stationarity, we require $|z^*| > 1$, that is, $|\phi_1| < 1$. If $|\phi_1| \geq 1$, the process is non-stationary: for example, if $\phi_1 = 1$ we get a random walk, where the variance grows indefinitely over time.

In summary, checking that the roots of the characteristic polynomial lie outside the unit circle is a mathematical way to guarantee that the process will not "blow up" or show unpredictable behaviors as time goes on, and that its statistical properties are stable and meaningful.

2.3 Properties of the AR(p)

Mean (Expectation) The unconditional mean (long-run average value) of a stationary AR(p) process with intercept is:

$$\mu = \mathbb{E}[X_t] = \frac{c}{1 - \sum_{i=1}^{p} \phi_i}. \tag{4}$$

This formula shows that the mean is determined by the constant c and the sum of AR coefficients. If $c = 0$, the process fluctuates around zero.

Variance The unconditional variance $\mathrm{Var}(X_t)$ measures the typical squared distance from the mean. For AR(1), it is:

$$\mathrm{Var}(X_t) = \frac{\sigma^2}{1 - \phi_1^2}. \tag{5}$$

For higher orders ($p > 1$), the variance is found by solving the Yule-Walker equations:

$$\gamma(k) = \sum_{i=1}^{p} \phi_i \gamma(k - i) \qquad (k \geq 1), \tag{6}$$

where $\gamma(k) = \mathrm{Cov}(X_t, X_{t-k})$.

Conditional Mean The conditional mean is the expected value of X_t given the previous p observations:

$$\mathbb{E}[X_t \mid X_{t-1}, \ldots, X_{t-p}] = c + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} \tag{7}$$

This is the best linear prediction of the next value using the past.

Conditional Variance Given the past, the conditional variance (uncertainty about X_t) is:

$$\mathrm{Var}(X_t \mid X_{t-1}, \ldots, X_{t-p}) = \mathrm{Var}(\epsilon_t) = \sigma^2 \tag{8}$$

because ϵ_t is independent from previous values.

Interpretation and Importance

- The unconditional mean and variance describe the long-term behavior: the average level and variability of the series.
- The conditional mean tells us how to predict the next value from the past (crucial for forecasting).
- The conditional variance tells us how much randomness remains after accounting for the past: for AR models, the future is uncertain only because of new, unpredictable shocks.

Example: If $X_t = 2 + 0.5X_{t-1} + \epsilon_t$, then the mean is $\mu = 2/(1 - 0.5) = 4$.

2.4 Properties of the MA(q) Model

The MA(q) model is always stationary for any parameter values.

Mean (Expectation) The unconditional mean of a stationary MA(q) model with mean parameter μ is:

$$\mathbb{E}[X_t] = \mu, \tag{9}$$

since $\mathbb{E}[\epsilon_t] = 0$.

Variance The unconditional variance is:

$$\text{Var}(X_t) = \sigma^2 \left(1 + \theta_1^2 + \cdots + \theta_q^2\right), \tag{10}$$

plus cross-terms if the MA parameters are not all zero. The variance is constant over time.

Conditional Mean The conditional mean, given the past, is simply the mean:

$$\mathbb{E}[X_t \mid \mathcal{F}_{t-1}] = \mu, \tag{11}$$

since future shocks are unpredictable and independent of the past.

Conditional Variance

$$\text{Var}(X_t \mid \mathcal{F}_{t-1}) = \text{Var}(X_t), \tag{12}$$

since all components contributing to X_t are either known (for $t-1, \ldots, t-q$) or unpredictable (ϵ_t), and past shocks cannot be observed.

Interpretation and Importance

- The MA(q) model captures short-term dependencies: a shock today influences only the next q periods.
- The conditional mean is the mean level, reflecting that future values cannot be predicted from the past in MA models.
- These models are useful for modeling processes with sudden, isolated shocks or measurement errors.

Example: If $X_t = 5 + \epsilon_t + 0.5\epsilon_{t-1}$, then the mean is 5, and the variance is $\sigma^2(1 + 0.5^2) = 1.25\sigma^2$.

2.5 Properties of the ARMA(p, q) Model with Intercept

The ARMA(p, q) model combines both autoregressive and moving average components:

$$X_t = c + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}. \tag{13}$$

Stationarity and Invertibility

- The process is stationary if the AR part's characteristic polynomial roots are outside the unit circle.
- The process is invertible (so past shocks can be uniquely recovered) if the MA polynomial has roots outside the unit circle.

Mean (Expectation) The unconditional mean of a stationary ARMA(p, q) model is:

$$\mu = \mathbb{E}[X_t] = \frac{c}{1 - \sum_{i=1}^{p} \phi_i}, \tag{14}$$

assuming the white noise has mean zero.

Variance and Autocovariance

- The unconditional variance and autocovariances are usually computed recursively via the ARMA equations and depend on both AR and MA parameters.
- The autocorrelation function of an ARMA(p, q) model typically displays both slow (AR-type) and rapid (MA-type) decay.

Conditional Mean The conditional mean given past values is:

$$\mathbb{E}[X_t \mid X_{t-1}, X_{t-2}, \ldots] = c + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q}, \tag{15}$$

where the ϵ_{t-i} are the estimated past innovations (often called "filtered" shocks).

Conditional Variance

$$\mathrm{Var}(X_t \mid \text{past}) = \sigma^2. \tag{16}$$

The uncertainty of the next value, given all past information, is due to the new shock ϵ_t.

Interpretation and Importance

- ARMA models are flexible and widely used in practice, as they capture both long-term dependencies and short-term shocks.
- The AR part models persistence (memory), while the MA part models the effect of recent, unpredictable shocks.
- Understanding these properties is fundamental for forecasting, model selection, and interpreting time series data.

Example: Suppose we model daily returns on a stock with $X_t = 0.1 + 0.6X_{t-1} + \epsilon_t + 0.4\epsilon_{t-1}$; then the mean is $0.1/(1 - 0.6) = 0.25$.

Relationship between AR(p) and MA(q) Models

Are AR and MA models equivalent?
No, in general an autoregressive model AR(p) and a moving average model MA(q) are not the same, and one cannot usually be written exactly as the other with a finite number of terms.

Infinite-order connection:

- Any **stationary AR(p)** model can be rewritten as an **MA(∞)** model: that is, as an infinite sum of present and past random shocks. The coefficients of this infinite expansion decay as we move further in the past.
- Any **invertible MA(q)** model can be rewritten as an **AR(∞)** model: that is, as an infinite sum of previous values of the process, with decaying weights.

Practical meaning:

- **AR(p):** The present depends directly on the past p values.
- **MA(q):** The present depends directly on the most recent q shocks (random disturbances).
- In practice, *any stationary, invertible time series process can be represented (at least approximately) by an ARMA(p, q) model*, which combines both effects.

Why is this important?

- This duality is fundamental in time series modeling: it shows that ARMA models are highly flexible and can capture a wide variety of real-world patterns.
- Understanding this connection helps us choose the right model and interpret time series data more effectively.

Summary Table:

Model	Infinite-order representation	Finite-order equivalence
AR(p)	MA(∞)	No, unless in special cases
MA(q)	AR(∞) if invertible	No, unless in special cases

3 Autoregressive Modeling (AR) of the 3-Month Treasury Bill Rate

In this section, we apply an autoregressive (AR) model to analyze the time series of U.S. 3-month Treasury bill rates. Stationarity is a fundamental assumption for the application of AR models, as non-stationary series can lead to spurious results and unreliable parameter estimates.

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import statsmodels.api as sm
from statsmodels.tsa.stattools import adfuller
from statsmodels.tsa.ar_model import AutoReg
from pandas_datareader import data as pdr
import datetime
import seaborn as sns
from google.colab import files

# === 1. Download DTB3 from FRED
start_date = datetime.datetime(1990, 1, 1)
end_date = datetime.datetime.today()

# Fetch data
dtb3 = pdr.DataReader("DTB3", "fred", start_date, end_date)
dtb3.dropna(inplace=True)

# Split into training and testing sets (last 2 years as test)
cutoff_date = end_date - datetime.timedelta(days=2 * 365)
train = dtb3[dtb3.index < cutoff_date]
test = dtb3[dtb3.index >= cutoff_date]

# === 2. AR(1) on original series
print("\n--- AR(1) on original series ---")
model_ar1_orig = AutoReg(train["DTB3"], lags=1, trend="n").fit()
print(model_ar1_orig.summary())

# === 3. ADF Test on training set
print("\n--- ADF Test on training set ---")
result = adfuller(train["DTB3"])
print("ADF Statistic:", result[0])
print("p-value:", result[1])

# === 4. Differencing if necessary
if result[1] < 0.05:
    print("Train series is likely stationary (ADF p < 0.05).")
    series_to_use = train["DTB3"]
    title = "3-Month Treasury Bill Rate (DTB3)"
    ylabel = "Rate"
else:
    print("Train series is non-stationary. Differencing applied.")
    series_to_use = train["DTB3"].diff().dropna()
    title = "First Difference of 3-Month U.S. Treasury Bill Rate"
    ylabel = "Differenced Rate (in %)"

# === 5. Plotting
sns.set(style="whitegrid", context="paper", font_scale=1.2)

fig, ax = plt.subplots(figsize=(10, 5))
ax.plot(series_to_use, linewidth=1.2, color='darkblue')
ax.set_title(title, fontsize=14)
ax.set_xlabel("Date", fontsize=12)
ax.set_ylabel(ylabel, fontsize=12)
ax.grid(True, linestyle='--', alpha=0.6)
plt.tight_layout()
```

```python
# === 6. Save and download
filename_resid = "diff_dtb3_plot.pdf"
plt.savefig(filename_resid, dpi=600, bbox_inches='tight')
plt.show()
files.download(filename_resid)

# === 7. ADF Test on final series
print("\n--- ADF Test on final series used (after differencing if applied) ---")
result = adfuller(series_to_use)
print("ADF Statistic:", result[0])
print("p-value:", result[1])
```

For this reason, after downloading daily data from January 1990 to the present, we assess the stationarity of the series using the Augmented Dickey-Fuller (ADF) test, where the number of lags is selected based on the Akaike Information Criterion (AIC). As shown in Table 1, the test fails to reject the null hypothesis of a unit root, indicating that the original series is non-stationary and thus unsuitable for direct AR modeling.

Table 1: ADF Test Results on the Original DTB3 Series

Test	**ADF Statistic**	**p-value**
Augmented Dickey-Fuller	-2.129	0.233

This conclusion is further supported by the estimation of an AR(1) model on the raw data. As reported in Table 2, the estimated autoregressive coefficient is approximately equal to one, with extremely narrow confidence bounds. This implies that the series behaves like a unit root process, exhibiting persistent, non-mean-reverting dynamics that invalidate the use of standard AR approaches in its original form.

Table 2: Estimation Results of the AR(1) Model on the Original DTB3 Series

Variable	**Coefficient**	**Std. Error**	**z-stat**	**p-value**	**95% CI**
DTB3_{t-1}	0.9998	0.0000	6803.270	0.000	[0.999, 1.000]

```python
# Fit AR(p) model to training data using BIC
best_bic = np.inf
best_model = None
best_lag = 0

for lag in range(1, 10):
    try:
        model = AutoReg(series_to_use, lags=lag, old_names=False).fit()
        if model.bic < best_bic:
            best_bic = model.bic
            best_model = model
            best_lag = lag
    except Exception as e:
        print(f"AR({lag}) failed:", e)

print(f"\nBest AR model: AR({best_lag}) with BIC = {best_bic:.2f}")
print(best_model.summary())
```

```
import matplotlib.pyplot as plt
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from google.colab import files

# === Residuals from best model
residuals = best_model.resid

# === ACF and PACF plot from lag 1
fig, ax = plt.subplots(2, 1, figsize=(10, 6))

# ACF
plot_acf(
    residuals,
    lags=10,
    ax=ax[0],
    zero=False,
    alpha=0.05,
    title="ACF of Residuals"
)
ax[0].set_ylim(-0.1, 0.1)

# PACF
plot_pacf(
    residuals,
    lags=10,
    ax=ax[1],
    method='ywm',
    zero=False,
    alpha=0.05,
    title="PACF of Residuals "
)
ax[1].set_ylim(-0.05, 0.1)

plt.tight_layout()

# === Save and download
filename_resid = "acf_pacf_ar_residuals.pdf"
plt.savefig(filename_resid, dpi=600, bbox_inches='tight')
plt.show()
files.download(filename_resid)
```

To address this issue, we apply a stationarizing transformation by taking the first differences of the series. Figure 1 displays the differenced series, and the corresponding ADF test results (Table 3) confirm that the transformed series is now statistically stationary.

Table 3: ADF Test Results on the Differenced DTB3 Series

Test	**ADF Statistic**	**p-value**
Augmented Dickey-Fuller	-12.190	1.29×10^{-22}

With a stationary series in place, we proceed to identify the most appropriate AR(p) specification. To do so, we estimate models with lag lengths ranging from 1 to 10 and select the one that minimizes the Bayesian Information Criterion (BIC), which balances model fit and parsimony. The results are reported in Table 4. The best model according to the BIC (**−28202.680**) is an AR(5), that is, an autoregressive model including five lagged terms. As shown in Table 4, all lag coefficients are statistically significant at the 1% level, whereas the intercept is not significantly different from zero.

This suggests that the dynamics of the differenced 3-month Treasury bill rate can be effectively captured by its own past values, with no need to include a constant term. The significance of multiple lags highlights the presence of short-term memory in the series.

To assess the adequacy of the AR(5) specification, we examine the residuals to verify whether they resemble white noise. Specifically, we plot the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the model residuals, as shown in Figure 2. The confidence bands correspond to the 95% significance level.

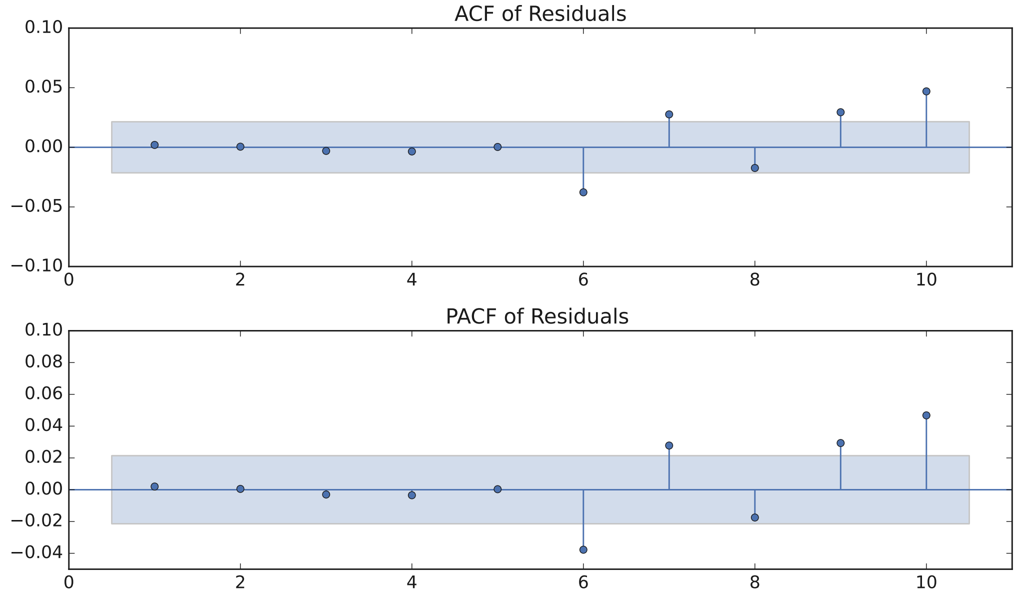

Figure 2: ACF and PACF of residuals from the AR(5) model estimated on the differenced DTB3 series.

The ACF and PACF plots confirm that the residuals from the AR(5) model are largely uncorrelated across lags, suggesting that the model has successfully captured the serial dependence in the differenced DTB3 series. All autocorrelations fall within the 95% confidence bands, except for a small spike in the ACF at lag 10. This isolated deviation may reflect random sampling variation rather than systematic misspecification. Overall, the residuals exhibit behavior close to white noise, supporting the adequacy of the AR(5) specification. We now turn to moving average (MA) models, which are particularly effective in modeling the impact of transitory shocks on time series dynamics.

4 Moving Average (MA) Modeling of the 3-Month Treasury Bill Rate

In this section, we empirically estimate a moving average (MA) model on the stationary DTB3 series. Given the series' short-term persistence identified in previous analyses, we explore MA specifications to assess whether past shocks alone can adequately capture the observed dynamics. We follow a procedure analogous to that used for the AR specification to estimate the MA model. It is important to note that, while a dedicated module

Table 4: Estimation Results of the AR(5) Model on the Stationary DTB3 Series. The model includes five autoregressive lags and was selected based on the lowest Bayesian Information Criterion. All lagged terms are statistically significant at the 1% level, while the constant is not.

Variable	**Coefficient**	**Std. Error**	**z-stat**	**p-value**	**95% CI**
const	-0.0003	0.0000	-0.546	0.585	[-0.001, 0.001]
DTB3_{t-1}	0.1260	0.0110	11.545	0.000	[0.105, 0.147]
DTB3_{t-2}	-0.0865	0.0110	-7.868	0.000	[-0.108, -0.065]
DTB3_{t-3}	-0.0871	0.0110	-7.923	0.000	[-0.109, -0.066]
DTB3_{t-4}	0.0368	0.0110	3.351	0.001	[0.015, 0.058]
DTB3_{t-5}	0.0547	0.0110	5.013	0.000	[0.033, 0.076]

(AutoReg) is available in statsmodels for the autoregressive case, this is not the case for MA models. In practice, MA models must be estimated using the more general ARIMA class, by setting both the autoregressive and differencing components to zero. That is, an MA(q) model is estimated by fitting an ARIMA(0, 0, q) specification. This approach enables consistent estimation and diagnostic testing within a unified framework for AR, MA, and ARMA processes.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from google.colab import files

# === 1. Fit MA(q) model by BIC selection
best_bic = np.inf
best_model = None
best_q = 0

for q in range(1, 10):
    try:
        model = ARIMA(series_to_use, order=(0, 0, q)).fit()
        if model.bic < best_bic:
            best_bic = model.bic
            best_model = model
            best_q = q
    except Exception as e:
        print(f"MA({q}) failed:", e)

print(f"\nBest MA model: MA({best_q}) with BIC = {best_bic:.2f}")
print(best_model.summary())


# === Residual analysis
residuals = best_model.resid

fig, ax = plt.subplots(2, 1, figsize=(10, 6))

plot_acf(residuals, lags=10, ax=ax[0], zero=False, alpha=0.05)
ax[0].set_title("ACF of Residuals (MA model)")
ax[0].set_ylim(-0.1, 0.1)
```

```python
plot_pacf(residuals, lags=10, ax=ax[1], zero=False, alpha=0.05, method='ywm')
ax[1].set_title("PACF of Residuals (MA model)")
ax[1].set_ylim(-0.05, 0.1)

plt.tight_layout()

# Save and download
filename_resid = "acf_pacf_ma_residuals.pdf"
plt.savefig(filename_resid, dpi=600, bbox_inches='tight')
plt.show()
files.download(filename_resid)
```

Table 5: Estimation Results of the MA(5) Model on the Stationary DTB3 Series. The model is estimated using the ARIMA(0,0,5) specification. All moving average terms are statistically significant at the 1% level, while the constant is not.

Variable	**Coefficient**	**Std. Error**	**z-stat**	**p-value**	**95% CI**
const	-0.0003	0.0006	-0.512	0.608	[-0.001, 0.001]
MA_1	0.1305	0.0031	41.505	0.000	[0.124, 0.137]
MA_2	-0.0727	0.0026	-28.382	0.000	[-0.078, -0.068]
MA_3	-0.1109	0.0035	-31.553	0.000	[-0.118, -0.104]
MA_4	0.0149	0.0030	4.941	0.000	[0.009, 0.021]
MA_5	0.0754	0.0039	19.497	0.000	[0.068, 0.083]

As shown in Table 5, the MA(5) model (BIC = $\mathbf{-28226.707}$) provides a statistically robust fit to the differenced DTB3 series. All five moving average coefficients are statistically significant at the 1% level, whereas the intercept remains not significantly different from zero. This supports the idea that the short-term dynamics of the series can be adequately captured by past shocks alone, without the need for a deterministic trend component. The resulting model achieves a lower BIC value than the best AR specification, indicating a superior in-sample fit. To further validate the model, we examine the residuals to verify the absence of serial correlation. Figure 3 displays the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the residuals obtained from the MA(5) model.

The plots indicate that the residuals are largely uncorrelated across lags, confirming that the MA(5) model has effectively captured the temporal dependence in the data. All autocorrelations lie within the 95% confidence bands, with only minor spikes at lags 9 and 10. These slight deviations are commonly regarded as random variation and do not indicate substantial model misspecification. Therefore, the residual diagnostics support the adequacy of the MA(5) specification and further reinforce its superior performance relative to the AR alternatives. Building upon the insights gained from both AR and MA models, we now turn to a more general specification that combines the strengths of both: the autoregressive moving average (ARMA) model.

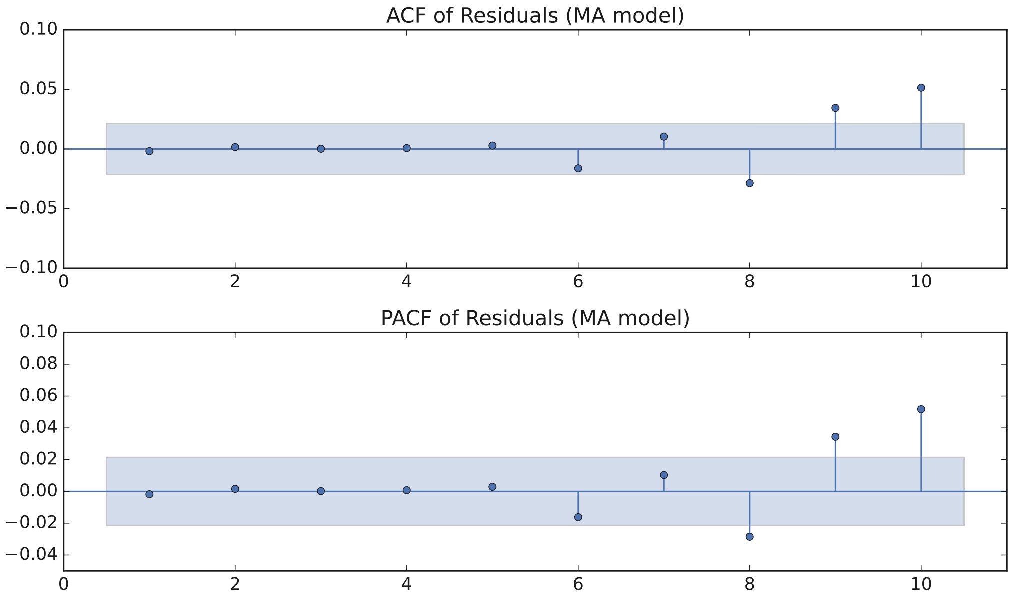

Figure 3: ACF and PACF of residuals from the MA(5) model estimated on the differenced DTB3 series.

5 ARMA Modeling of the 3-Month Treasury Bill Rate

We now turn to the empirical estimation of ARMA models on the stationary DTB3 series. After analyzing autoregressive and moving average specifications separately, this section investigates whether a combined approach offers a better representation of the underlying dynamics.

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from statsmodels.tsa.arima.model import ARIMA
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from google.colab import files

# === 1. Fit ARMA(p,q) model by BIC selection
best_bic = np.inf
best_model = None
best_p = 0
best_q = 0

for p in range(1, 6):
    for q in range(1, 6):
        try:
            model = ARIMA(series_to_use, order=(p, 0, q)).fit()
            if model.bic < best_bic:
                best_bic = model.bic
                best_model = model
                best_p = p
                best_q = q
        except Exception as e:
            print(f"ARMA({p},{q}) failed:", e)
```

```
print(f"\nBest ARMA model: ARMA({best_p},{best_q}) with BIC = {best_bic:.2f}")
print(best_model.summary())

# === 2. Residual analysis
residuals = best_model.resid

fig, ax = plt.subplots(2, 1, figsize=(10, 6))

plot_acf(residuals, lags=10, ax=ax[0], zero=False, alpha=0.05)
ax[0].set_title("ACF of Residuals (ARMA model)")
ax[0].set_ylim(-0.1, 0.1)

plot_pacf(residuals, lags=10, ax=ax[1], zero=False, alpha=0.05, method='ywm')
ax[1].set_title("PACF of Residuals (ARMA model)")
ax[1].set_ylim(-0.05, 0.1)

plt.tight_layout()

# === 3. Save and download plot
filename_resid = "acf_pacf_arma_residuals.pdf"
plt.savefig(filename_resid, dpi=600, bbox_inches='tight')
plt.show()
files.download(filename_resid)
```

Following the same strategy adopted for the AR and MA models, we estimate a grid of ARMA(p, q) models with lag orders ranging from 1 to 10. The optimal specification is selected according to the Bayesian Information Criterion (BIC), which balances model fit and complexity. All models are estimated using the `ARIMA` class in `statsmodels`, with the differencing parameter set to zero to reflect the pre-transformed (stationary) nature of the input series.

Table 6: Estimation results for the ARMA(3,2) model fitted to the differenced DTB3 series.

Parameter	**Coefficient**	**Std. Error**	**z-Statistic**	**p-Value**	**[95% CI]**
Constant	-0.0009	0.001	-1.621	0.105	[-0.002, 0.000]
AR(1)	0.5381	0.008	71.066	0.000	[0.523, 0.553]
AR(2)	-0.9726	0.006	-151.989	0.000	[-0.985, -0.960]
AR(3)	0.0757	0.004	20.722	0.000	[0.069, 0.083]
MA(1)	-0.4130	0.007	-55.720	0.000	[-0.428, -0.399]
MA(2)	0.8994	0.006	138.693	0.000	[0.887, 0.912]
σ^2	0.0020	6.28e-06	310.791	0.000	[0.002, 0.002]

As shown in Table 6, the ARMA(3,2) model offers an effective characterization of the dynamics in the differenced DTB3 series. All autoregressive and moving average coefficients are highly statistically significant at the 1% level, indicating that both past values and past shocks contribute meaningfully to the process. The intercept is not statistically different from zero.

The model outperforms both the best AR and MA specifications in terms of Bayesian Information Criterion (BIC), achieving a value of $\mathbf{-28280.05}$, the lowest among all tested models. This suggests that incorporating both AR and MA components provides a better in-sample fit without excessive complexity. We proceed by evaluating the residuals to ensure that the model adequately captures the serial dependence structure.

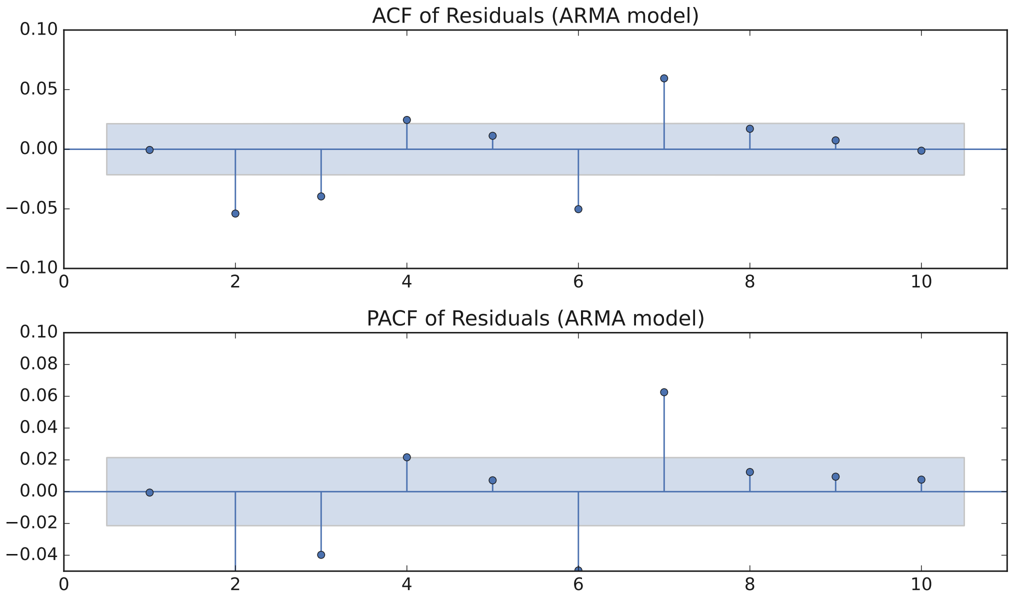

Figure 4: ACF and PACF of residuals from the ARMA(3,2) model estimated on the differenced DTB3 series.

Figure 4 displays the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the residuals obtained from the ARMA(3,2) model. The residuals appear approximately white noise, as the autocorrelation and partial autocorrelation values lie well within the 95% confidence bounds across all lags. Only a modest spike is observed at lag 7 in the PACF, but it remains within the bounds and does not indicate significant serial correlation.

These diagnostic plots confirm that the ARMA(3,2) specification has effectively captured the temporal dependence in the data, leaving no substantial structure in the residuals. Combined with the model's strong BIC performance and statistically significant parameters, this reinforces the suitability of the ARMA(3,2) model for describing the short-run dynamics of the differenced 3-month Treasury Bill rate.

6 Conclusion

AR, MA, and ARMA models provide a powerful and flexible framework for analyzing the dynamics of financial time series. Their mathematical simplicity, interpretability, and forecasting ability make them a cornerstone in econometric modeling and financial practice.

Through the empirical analysis of the 3-month U.S. Treasury Bill rate, we have demonstrated how these models can be used to identify and capture both persistent and transitory components in time series data. The process of stationarizing the series, selecting appropriate lag orders using information criteria such as BIC, and conducting residual diagnostics provides a structured and rigorous methodology for model development and evaluation.

While AR models are effective in capturing autoregressive dependence and MA models excel in modeling the effects of random shocks, the ARMA framework offers a balanced combination of both. In practice, the selection of the appropriate model should be guided not only by statistical fit but also by the economic or financial context and interpretability.

These classical linear models also serve as a foundational stepping stone toward more advanced approaches, such as ARIMA (for integrated processes), SARIMA (seasonal effects), or nonlinear and volatility models like GARCH. However, even in the presence of more complex alternatives, AR, MA, and ARMA models remain highly relevant for their clarity, robustness, and ease of implementation.

References

[1] Yule, G. U. (1927). On a method of investigating periodicities in disturbed series, with special reference to Wolfer's sunspot numbers. *Philosophical Transactions of the Royal Society of London. Series A*, 226, 267–298.

[2] Slutsky, E. (1937). The Summation of Random Causes as the Source of Cyclic Processes. *Econometrica*, 5(2), 105–146.

[3] Box, G. E. P., & Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day.