# DATA 201: Time Series Analysis
# Linear Regression Model
# Lecture 11: The Auto-Regressive (AR) Model

Lulu Wang

Data Analytics
Dickinson College

2/28/2025

## The Auto-Regressive (AR) Model

## Estimation

## Order selection

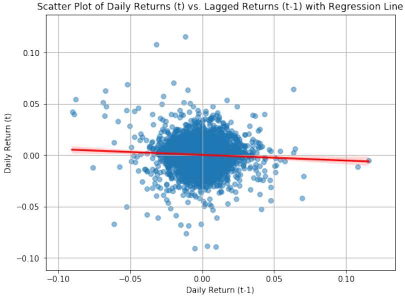
# Auto-Regressive (AR) model

- The AutoRegressive model of order 1, denoted by **AR(1)**, is given by:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \epsilon_t$$

where $\beta_0$ and $\beta_1$ are parameters to be estimated and $\epsilon_t$ has mean zero and variance $\sigma_\epsilon^2$.

- A scatter plot of $Y_{t-1}$ and $Y_t$ shows little dependence between the current and past value.
- The red line represents $E(Y_t|Y_{t-1}) = \beta_0 + \beta_1 Y_{t-1}$.



# Auto-Regressive (AR) Model

- In general, an **AR(p)** model is defined as:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} + \epsilon_t$$

where:

- $\beta_0, \beta_1, \ldots, \beta_p$ are parameters to be estimated.
- $\epsilon_t$ is a white noise error term with mean zero and variance $\sigma_\epsilon^2$.

# Conditional Expectation in AR(p) Model

**Property 1: Conditional Expectation**

$$E(Y_t|Y_{t-1},\dots,Y_{t-p}) = \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p}$$

It can be interpreted as a **forecast** since only past information is used to produce an expectation of the variable today.

**Property 2: Unconditional Expectation**

$$E(Y_t) = \frac{\beta_0}{1 - \sum_{j=1}^{p} \beta_j}$$

since:

$$E(Y_t) = \beta_0 + \beta_1 E(Y_{t-1}) + \cdots + \beta_p E(Y_{t-p}) + E(\epsilon_t)$$

Assuming that $E(Y_t) = E(Y_{t-k})$ for all values of $k$.
It represents the long-run mean of $Y_t$

# Conditional Expectation in AR(p) Model

**Proof of Property 1: Conditional Expectation**

The AR(p) model is given by:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \cdots + \beta_p Y_{t-p} + \varepsilon_t \tag{1}$$

Taking the expectation conditional on past values:

$$E(Y_t | Y_{t-1}, \ldots, Y_{t-p}) = E\left(\beta_0 + \sum_{j=1}^{p} \beta_j Y_{t-j} + \varepsilon_t \middle| Y_{t-1}, \ldots, Y_{t-p}\right) \tag{2}$$

Since $E(\varepsilon_t | Y_{t-1}, \ldots, Y_{t-p}) = 0$, we obtain:

$$E(Y_t | Y_{t-1}, \ldots, Y_{t-p}) = \beta_0 + \sum_{j=1}^{p} \beta_j Y_{t-j} \tag{3}$$

# Variance of AR(p) Model

**Property 3: Variance**

$$Var(Y_t) = \frac{\sigma_\epsilon^2}{1 - \sum_{j=1}^{p} \beta_j^2}$$

since:

$$Var(Y_t) = \beta_1^2 Var(Y_{t-1}) + \cdots + \beta_p^2 Var(Y_{t-p}) + Var(\epsilon_t)$$

$$= \left( \sum_{j=1}^{p} \beta_j^2 \right) Var(Y_t) + \sigma_\epsilon^2$$

This shows how past values contribute to the variance of the current value.

# Persistence and Mean Reversion in AR(p)

**Property 4: Persistence and Mean Reversion**

$$\sum_{i=1}^{p} \beta_i$$

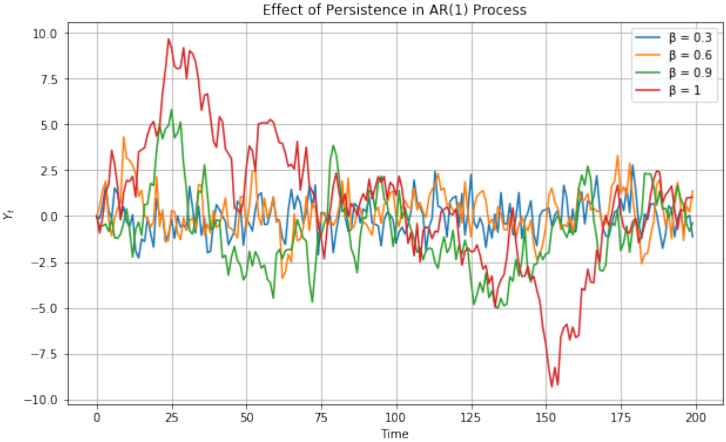
measures the **persistence** of the time series. Persistence can also be interpreted as **mean reversion**, which represents how quickly a time series reverts back to its mean.

- **Low persistence** $\Rightarrow$ quick mean reversion.
- **High persistence** $\Rightarrow$ slow mean reversion.

**Key Insight:** The closer the sum of $\beta_i$ is to 1, the more persistent the process, meaning past values strongly influence future values.

# Persistence and Mean Reversion in AR(1)

The plot below shows the **effect of persistence** in an **AR(1) process** with different values of $\beta$:



# Persistence and Mean Reversion in AR(1)

**Low Persistence** ($\beta = 0.3, 0.6$)

- The process exhibits **quick mean reversion**, meaning that shocks **fade away quickly**.
- The series remains **close to zero**, with less variability over time.

**High Persistence** ($\beta = 0.9$)

- The process takes **longer to revert to the mean**, meaning past values strongly influence future values.
- Large deviations **persist longer** before dying out.

**Unit Root** ($\beta = 1.0$)

- This is a **random walk**, meaning the process **never mean-reverts**.
- Variance **grows indefinitely**, and the series **does not stabilize**.

# Estimation of AR models

- Auto-regressive models AR(P) can be estimated consistently using OLS.
- More general time series models (e.g., MA, ARMA) need to be estimated by *Maximum Likelihood* (ML).
- There are several functions to estimate time series models that allow choosing the estimation method.
  - In practice, we can go with the default option since for moderate and large sample sizes, the results are nearly identical.
- We will discuss three functions (first two today):
  - Using `statsmodels.OLS()`: This method applies Ordinary Least Squares (OLS) estimation. **Make sure you use the Newey-West method!**
  - Using `statsmodels.tsa.ar_model.AutoReg()`:It uses Conditional Maximum Likelihood which is Ordinary Least Squares (OLS).
  - Using `statsmodels.tsa.arima.model.ARIMA()`: This function estimates AR, MA, and ARIMA models using ML.

# How Correlated Observations Lead to Dependent Errors

Consider a simple first-order autoregressive model $AR(1)$:

$$y_t = \beta y_{t-1} + \varepsilon_t \tag{4}$$

where:

- $y_t$ is the observation at time $t$,
- $\beta$ is the autoregressive coefficient, where $|\beta| < 1$ for stationarity,
- $\varepsilon_t$ is a white noise error term.

By recursively substituting:

$$y_{t-1} = \beta y_{t-2} + \varepsilon_{t-1} \tag{5}$$

$$y_t = \beta(\beta y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t = \beta^2 y_{t-2} + \beta \varepsilon_{t-1} + \varepsilon_t \tag{6}$$

# Autocovariance of the Errors

The error term at time $t$ is:

$$\varepsilon_t = y_t - \beta y_{t-1} \tag{7}$$

Similarly, the error at $t-1$ is:

$$\varepsilon_{t-1} = y_{t-1} - \beta y_{t-2} \tag{8}$$

Computing the covariance:

$$\mathrm{Cov}(\varepsilon_t, \varepsilon_{t-1}) = \mathrm{Cov}(y_t - \beta y_{t-1}, y_{t-1} - \beta y_{t-2}) \tag{9}$$

Expanding using linearity of covariance:

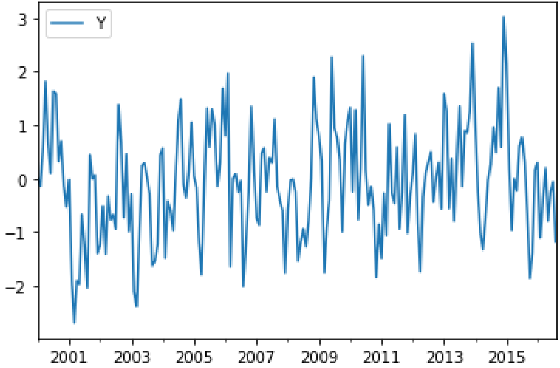
$$\begin{aligned}\mathrm{Cov}(\varepsilon_t, \varepsilon_{t-1}) = {} & \mathrm{Cov}(y_t, y_{t-1}) - \beta \mathrm{Cov}(y_t, y_{t-2}) \\ & - \beta \mathrm{Cov}(y_{t-1}, y_{t-1}) + \beta^2 \mathrm{Cov}(y_{t-1}, y_{t-2})\end{aligned} \tag{10}$$

This results in a non-zero covariance, meaning the errors are correlated over time.

# AR estimated by `OLS()`

- We will simulate time series data.

$$Y_t = 0.5 Y_{t-1} + \epsilon_t$$



# AR estimated by `OLS()`

- Estimation summary of

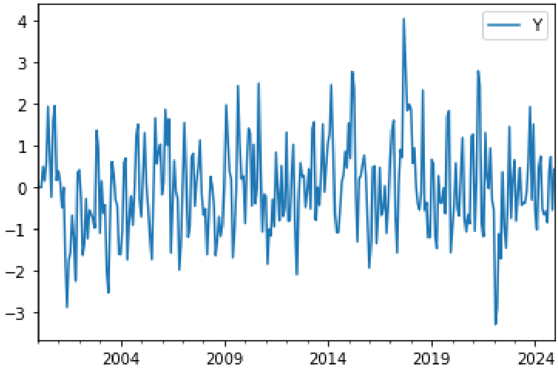
$$Y_t = 0.5 Y_{t-1} + \epsilon_t$$

```
=== OLS Results (HAC/ Newey-West Standard Errors) ===
                            OLS Regression Results
==============================================================================
Dep. Variable:                      Y   R-squared:                       0.170
Model:                            OLS   Adj. R-squared:                  0.168
Method:                 Least Squares   F-statistic:                     60.83
Date:                Thu, 27 Feb 2025   Prob (F-statistic):           1.07e-13
Time:                        18:59:17   Log-Likelihood:                -425.08
No. Observations:                 298   AIC:                             854.2
Df Residuals:                     296   BIC:                             861.5
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0145      0.059     -0.247      0.805      -0.130       0.101
Y_lag1         0.4130      0.053      7.799      0.000       0.309       0.517
==============================================================================
Omnibus:                        3.409   Durbin-Watson:                   1.858
Prob(Omnibus):                  0.182   Jarque-Bera (JB):                3.090
Skew:                           0.217   Prob(JB):                        0.213
Kurtosis:                       3.247   Cond. No.                         1.11
==============================================================================
```

# AR estimated by `AutoReg()`

- Example: estimating a AR(3) model:

$$Y_t = 0.6Y_{t-1} - 0.3Y_{t-2} + 0.2Y_{t-3} + \epsilon_t$$



# AR estimated by `AutoReg()`

- Estimation summary of

$$Y_t = 0.6Y_{t-1} - 0.3Y_{t-2} + 0.2Y_{t-3} + \epsilon_t$$

```
                            AutoReg Model Results
==============================================================================
Dep. Variable:                      Y   No. Observations:                  300
Model:                     AutoReg(3)   Log Likelihood                -415.822
Method:               Conditional MLE   S.D. of innovations              0.981
Date:                Mon, 24 Feb 2025   AIC                            841.644
Time:                        18:49:08   BIC                            860.113
Sample:                    04-30-2000   HQIC                           849.038
                         - 12-31-2024
==============================================================================
                 coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------
const         -0.0137      0.057     -0.240      0.810      -0.125       0.098
Y.L1           0.5116      0.057      8.925      0.000       0.399       0.624
Y.L2          -0.2497      0.063     -3.962      0.000      -0.373      -0.126
Y.L3           0.1625      0.057      2.829      0.005       0.050       0.275
                                    Roots
=============================================================================
                  Real          Imaginary           Modulus         Frequency
-----------------------------------------------------------------------------
AR.1            1.7486           -0.0000j            1.7486           -0.0000
AR.2           -0.1060           -1.8730j            1.8760           -0.2590
AR.3           -0.1060           +1.8730j            1.8760            0.2590
-----------------------------------------------------------------------------
```

# Order selection

- **How do we choose the order $p$ of the AR model?**
- The problem is the same as choosing the best LRM:
  - we use adjusted $R^2$ to penalize models that include more variables
  - select the model with largest adjusted $R^2$
- We do the same for time series models, but instead of using adjusted $R^2$ we use:
  1. *Akaike Information Criterion (AIC)* defined as

$$\log(RSS/T) + 2 \cdot \frac{(1+p)}{T}$$

  2. *Bayesian Information Criterion (BIC)* defined as

$$\log(RSS/T) + \log(T) \cdot \frac{(1+p)}{T}$$

- For AIC and BIC the best model has the *lowest* value of the criterion

# Order selection (Cont.)

- This can be achieved in Python using the `AutoReg` function from `statsmodels` by selecting the model with the lowest AIC or BIC.
- However, a more convenient approach is to use `ar_select_order()`, which automatically determines the optimal lag order based on AIC or BIC.