Training Manual

Advanced Forecasting Methods for Agricultural Data Analysis

Compiled and Edited by:

Dr. Md. Mosharraf Uddin Molla, MD (AERS), BARC
Dr. Md. Shofiqul Islam, PSO (AERS), BARC

February 2025

Agricultural Economics and Rural Sociology Division
Bangladesh Agricultural Research Council (BARC)

Training Manual

Advanced Forecasting Methods for Agricultural Data Analysis

23-27 February, 2025
Venue: Computer Training Lab, AIC Building, BARC

Dr. Md. Mosharraf Uddin Molla
Course Director

Dr. Md. Shofiqul Islam
Course Coordinator

The Training Module is Designed for Scientists of NARS Institutes

Organized by

Agricultural Economics and Rural Sociology Division
Bangladesh Agricultural Research Council (BARC)

Program Schedule

Day/Date	Time	Topic/Event	Resource person
23-02-25 Sunday	9:00-9:15	Registration	--
	9:15-09:35	Pre-Evaluation	
	9:35-10:15	Opening	--
	10:15-10.30	Tea Break	--
	10:30-11:20	Importance of forecasting in agricultural research	Dr. Nazmun Nahar Karim, EC, BARC
	11:20-12:10	Socio-economic research priorities in agriculture	Dr. Md. Mosharraf Uddin Molla
	12:10-1:00	Introduction of deterministic approach for future forecasting of agricultural data	Dr. S.M. Sayem, BAU
	1:00-2:00	Lunch	
	2:00-3:00	Hands-on exercise of deterministic approach for forecasting food consumption using STATA software	Dr. S.M. Sayem, BAU
	3:00-4:00	Regression with ARIMA error for forecasting	Dr. S.M. Sayem, BAU
24-02-25 Monday	9.30-10.30	Hands-on exercise of regression with ARIMA error for forecasting agricultural production	Dr. S.M. Sayem, BAU
	10.30-11:00	Tea break	
	11:00-12:00	Practicing SARIMA modeling for forecasting agricultural responses using STATA software	Dr. S.M. Sayem, BAU
	12:00-1.00	Basic concepts of dynamic regression models	Dr. S.M. Sayem, BAU
	1.00-2.00	Lunch	
	2.00-3.30	Hands-on exercise of dynamic regression models using STATA software	Dr. S.M. Sayem, BAU
	3.30-4.30	Testing of co-integration and causality	Dr. S.M. Sayem, BAU
25-02-25 Tuesday	9.30-10.30	Introduction of error correction mechanism (ECM) in agricultural research	Dr. S.M. Sayem, BAU
	10.30-11:00	Tea break	
	11:00-12:00	Application of error correction mechanism and co-integration in agricultural research	Dr. S.M. Sayem, BAU
	12:00-1.00	Basic concepts of intervention analysis	Dr. S.M. Sayem, BAU
	1.00-2.00	Lunch	
	2.00-3.30	Hands-on exercise of intervention analysis using STATA software	Dr. S.M. Sayem, BAU
	3.30-4.30	Key concepts and applications of the ARCH model for forecasting agricultural data	Dr. S.M. Sayem, BAU
26-02-25 Wednesday	9.30-10.30	Introduction of GARCH model for forecasting agricultural data	Dr. S.M. Sayem, BAU
	10.30-11:00	Tea break	

Day/Date	Time	Topic/Event	Resource person
	11:00-12:00	Hands-on exercise of ARCH and GARCH model for forecasting agricultural data using STATA software	Dr. S.M. Sayem, BAU
	12:00-1.00	Key concepts of different non-linear econometric models in agricultural research	Dr. S.M. Sayem, BAU
	1.00-2.00	Lunch	
	2.00-3.30	Documentation of the projection of demand and supply of agricultural crops in Bangladesh	Dr. Md. Mosharraf Uddin Molla
	3.30-4.30	Introduction of multivariate regression analysis: techniques and applications	Dr. S.M. Sayem, BAU
27-02-25 Thursday	9.30-10.30	Hands-on exercise of multivariate regression analysis using STATA software	Dr. S.M. Sayem, BAU
	10.30-11:00	Tea break	
	11:00-12:00	Basic concepts of multivariate autoregressive model for forecasting agro-economic data	Dr. S.M. Sayem, BAU
	12:00-1.00	Practicing multivariate autoregressive model for forecasting agro-economic data using STATA software	Dr. S.M. Sayem, BAU
	1.00-2.00	Lunch	
	2.00-3.00	Development of adaptive expectation model for prediction	Dr. Md. Abdus Salam
	3.00-4.00	Research scope in socio-economic perspectives: Advancements and challenges in forecasting	Dr. Md. Shofiqul Islam
	4.00-4.30	Post-Evaluation	
	4.00-5.00	Closing & certificate awarding ceremony	

Resource Persons

1. Dr. Md. Mosharraf Uddin Molla, Member Director, AERS Division, BARC
2. Professor Dr. Sheikh Mohammad Sayem, Dept. of Agricultural and Applied Statistics, BAU, Mymensingh
3. Dr. Md. Abdus Salam, Principal Scientific Officer, AERS Division, BARC
4. Dr. Md. Shofiqul Islam, Principal Scientific Officer, AERS Division, BARC

Contents

Introduction of deterministic approach for future forecasting of agricultural data

Prof. Dr. Sheikh Mohammad Sayem

Agricultural data forecasting is critical for effective planning, resource management, and decision-making in agriculture. Accurate forecasts can help farmers, policymakers, and stakeholders plan for future demands, manage supply chains, optimize crop yields, and address potential risks (such as climate change, market fluctuations, and pest outbreaks).

🕮 Deterministic vs. Stochastic Models in Forecasting

Stochastic Models: These models account for randomness or unpredictability. They rely on probability distributions and incorporate uncertainty. In agriculture, these models could include randomness due to weather variations, market price fluctuations, or pest infestations.

Examples of stochastic models include:

- ARIMA (AutoRegressive Integrated Moving Average) model
- GARCH (Generalized Autoregressive Conditional Heteroscedasticity) model

Deterministic Models: These models, on the other hand, predict future outcomes with certainty, given the input data and parameters. They assume that future events are entirely predictable from past data and there is no randomness involved. In agriculture, deterministic models are often used when the goal is to capture long-term trends, cyclic behaviors, or seasonality.

Examples of deterministic models include:

- Linear regression models
- Exponential smoothing
- Growth models

📖 Concepts in the Deterministic Approach

The deterministic forecasting approach focuses on identifying **patterns** or **relationships** in the historical data and using them to predict future outcomes. The key elements of this approach are:

1. **Trend Analysis**
    - In agriculture, trends represent the general movement or direction in data over time. For example, an upward trend in crop yields due to improved agricultural practices or a downward trend due to soil degradation can be modelled deterministically.
    - Common methods for modelling trends include **linear regression** and **polynomial regression**.
2. **Seasonality**
    - Agricultural data often exhibits seasonal patterns. For instance, crop yields might peak during harvest season and dip during non-growing periods.
    - **Seasonal decomposition** methods can be used to model these repetitive patterns.
3. **Cyclic Patterns**
    - While seasonality refers to regular, predictable fluctuations, **cycles** are longer-term, often irregular patterns influenced by external factors (e.g., market cycles or economic booms and busts).
    - Models like the **Hodrick-Prescott filter** or **Fourier analysis** can help isolate and predict these cycles.
4. **Growth Models**
    - Agricultural growth often follows patterns that can be described by deterministic growth functions. For example, the growth of crops over time may follow a **logistic growth model** that accounts for limited resources (such as space and nutrients).
    - Common deterministic growth models include:
        - ✓ **Exponential Growth Model**: Useful for modelling initial growth phases where the growth rate is constant.
        - ✓ **Logistic Growth Model**: Used when the growth rate slows as it approaches a carrying capacity.

📖 Deterministic Methods for Agricultural Forecasting

a. Linear Regression

Linear regression is one of the simplest and most widely used deterministic methods in forecasting. It models the relationship between an independent variable (e.g., time) and a dependent variable (e.g., crop yield or price) using a straight line.

For example:

$$Y_t = \beta_0 + \beta_1 t + \epsilon_t$$

WhereY_t is the dependent variable (e.g., crop yield at time t),t is the time variable (e.g., years or months),β_0 is the intercept (constant term),β_1 is the slope (rate of change), andϵ_t represents the residual error.

b. Exponential Smoothing

Exponential smoothing is a forecasting method that gives more weight to recent observations, making it suitable for situations where recent data is a better predictor of the future than older data.

The method uses the formula:

$$\hat{Y}_{t+1} = \alpha Y_t + (1 - \alpha)\hat{Y}_t$$

Where:

$\hat{Y}_{t+1}$is the forecast for the next time period,

$Y_t$is the observed value at time t,

$\hat{Y}_t$is the predicted value at time t, and

$\alpha$is the smoothing parameter $(0 < \alpha < 1)$.

c. Growth Models (Exponential Models)

Growth models can be useful when modelling the growth of agricultural products over time. Two common types are:

- **Exponential Growth**: Assumes that growth occurs at a constant rate over time and can be described by the equation:

  $$Y_t = Y_0 e^{rt}$$

  Where $Y_t$is the value at time t,$Y_0$is the initial value, r is the growth rate, and$e$ is the base of the natural logarithm.

d. Time Series Decomposition

Time series decomposition involves breaking down a time series into its components: trend, seasonality, and residuals. In agricultural forecasting, this technique is useful for understanding the underlying patterns in the data and forecasting future outcomes. The decomposition model is generally expressed as:

$$Y_t = T_t + S_t + R_t$$

Where:

Y_t is the observed data at time t,

T_t is the trend component,

S_t is the seasonal component, and

R_t is the residual (random error) component.

Applications of Deterministic Forecasting in Agriculture

1. **Crop Yield Forecasting**: Deterministic models can be used to predict crop yields based on historical data, taking into account seasonal effects and long-term trends.
2. **Price Prediction**: Agricultural commodities often exhibit seasonal and trend patterns. Deterministic models can be applied to predict future prices of crops like wheat, maize, and soybeans.
3. **Irrigation and Water Management**: Deterministic models can forecast water usage based on crop growth models, helping optimize irrigation schedules and water conservation.
4. **Climate Impact Analysis**: Forecasting the impact of climate change on agriculture can involve deterministic models to predict future yield reductions or shifts in planting seasons based on long-term temperature and precipitation trends.

Advantages of the Deterministic Approach

- ✓ **Simplicity**: Deterministic models are often simpler to implement and understand compared to stochastic models.
- ✓ **Predictability**: These models are useful when you are confident that the underlying data exhibits predictable patterns without much randomness.
- ✓ **No Need for Complex Probability Structures**: Unlike stochastic models, deterministic approaches don’t require knowledge of probability distributions, making them easier to implement in scenarios with limited data.

Challenges and Limitations

- ⌦ **Lack of Uncertainty**: The deterministic approach ignores the inherent uncertainty and randomness present in agricultural systems (e.g., unpredictable weather events).
- ⌦ **Overfitting**: If the model is too simplistic, it may fail to capture more complex dynamics, leading to inaccurate predictions.
- ⌦ **Assumption of Continuity**: These models assume that future values are determined by past data in a continuous way, which might not always be true for agricultural systems that are impacted by sudden external factors.

Conclusion

The deterministic approach to forecasting agricultural data offers valuable insights for predicting future outcomes based on historical patterns. While it has its limitations particularly in dealing with uncertainty it remains a useful tool in scenarios where trends, seasonality, and growth patterns are the primary focus. By using methods like linear regression, exponential smoothing, and growth models, agricultural stakeholders can make more informed decisions about crop management, pricing, and resource allocation.

References

1. Box, G.E.P., Jenkins, G.M., & Reinsel, G.C. (2015). *Time Series Analysis: Forecasting and Control*. 5th edition.
2. Gardner, B.D., & McKenzie, E.J. (1985). *Exponential Smoothing: The State of the Art*. International Journal of Forecasting.
3. Buis, M.L. (2010). *Deterministic and Stochastic Models in Forecasting Agricultural Data*. Springer.

Hands-on exercise of deterministic approach for forecasting food consumption using STATA software

Prof. Dr. Sheikh Mohammad Sayem

Problem statement: This data involves analyzing and forecasting monthly food consumption using time series techniques. The dataset spans 15 years (2010–2024) and includes key economic and demographic factors that influence food consumption patterns. The primary goal is to understand trends, detect seasonality, and make future projections using deterministic approach namely; smoothing methods such as Exponential Smoothing, Holt's Method, and Holt-Winters Seasonal Smoothing.

This artificial dataset consists of **180 monthly observations**, capturing the following variables:

Variable	Description
date	Monthly time variable (formatted correctly for time-series analysis).
year	Extracted year from the date variable.
month	Extracted month from the date variable.
population	Simulated population growth with some randomness.
income	Simulated rising income levels over time.
food_price	Cyclical variations in food prices over time.
food_consumption	Dependent variable representing monthly food consumption, affected by other factors.

The food consumption variable is influenced by income, population, and food prices, with an added seasonal component to reflect real-world fluctuations.

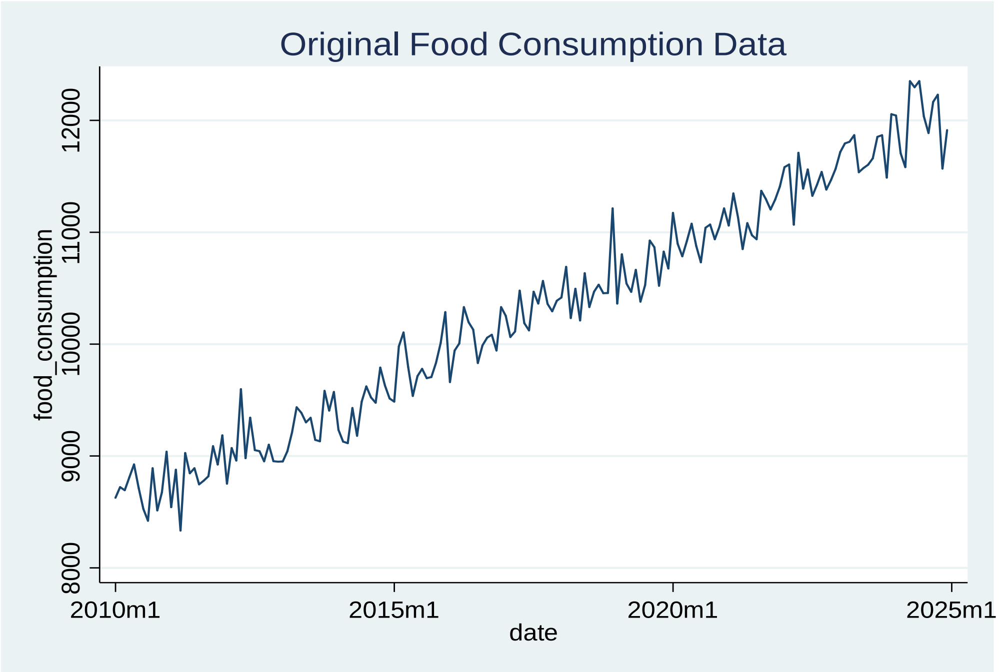
Requirements:

1. Understand time-series data structure and format data correctly for forecasting.
2. Visualize historical trends and detect volatility or seasonality in food consumption.
3. Apply smoothing techniques, including:
   - → Single Exponential Smoothing (for trend estimation).
   - → Double Exponential Smoothing (Holt's Method) (for trend correction).
   - → Holt-Winters Smoothing (for trend and seasonality modeling).

→ Moving Average Smoothing (for short-term fluctuations).

4. Evaluate forecasting models using Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).
5. Forecast future food consumption for the next 12 months (2025) using the best-fit model.

Visualizing the Data: Plot the original **food consumption trend** over time-



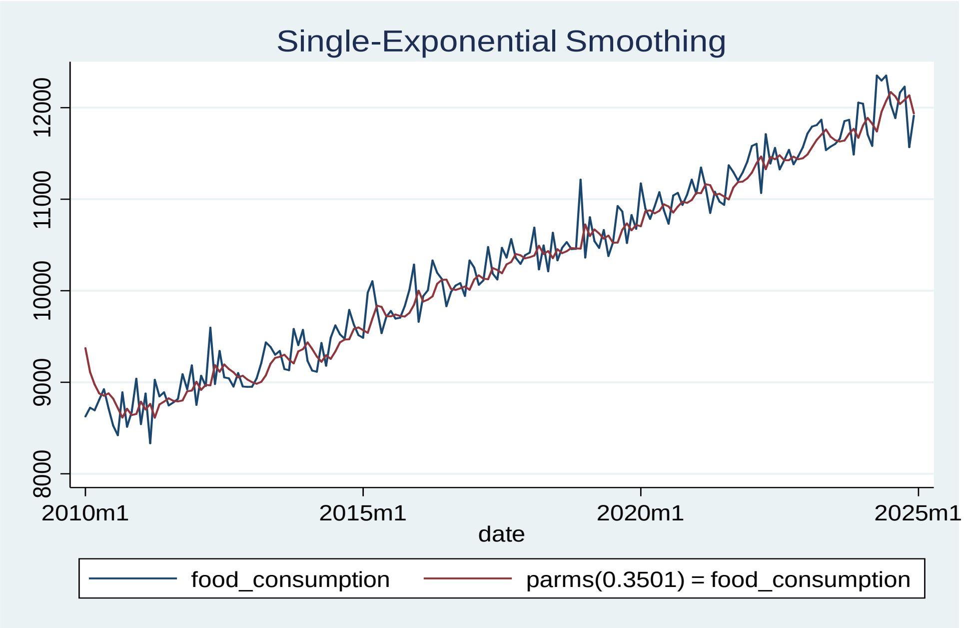
A fluctuating but upward-trending time series exist due to income and population growth, with clear seasonality.

Smoothing Techniques

We apply various **time-series smoothing methods** to identify trends and reduce fluctuations.

1. Single Exponential Smoothing

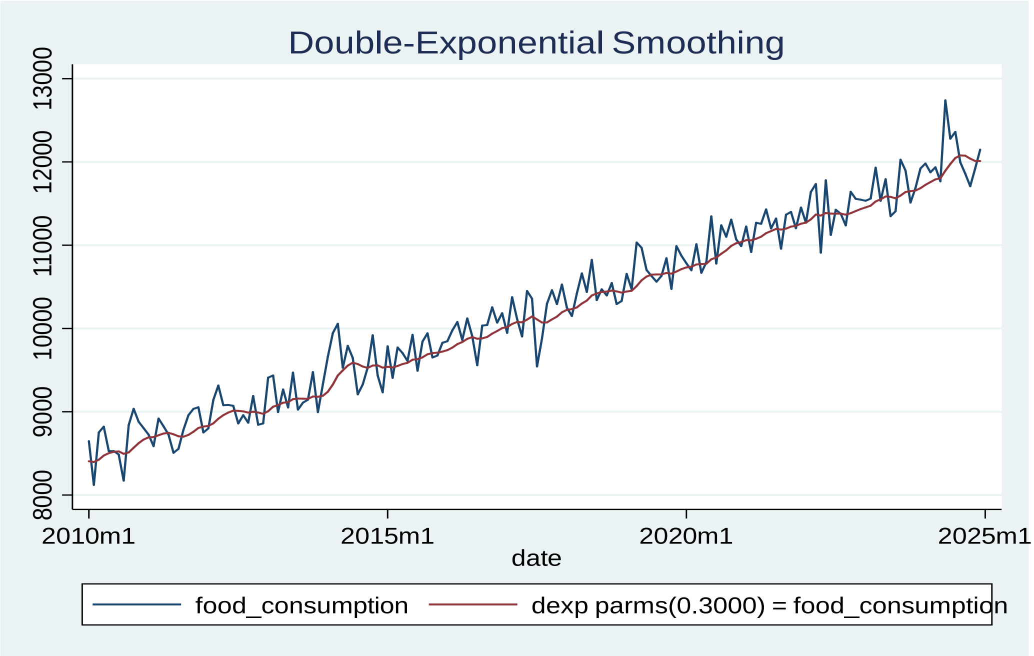
computing optimal exponential coefficient (0,1)
optimal exponential coefficient = 0.3415
sum-of-squared residuals = 12736404
root mean squared error = 266.00338



A **smoothed version of food consumption observes** reacting slower to fluctuations.

2. Double Exponential Smoothing (Holt's Method)

double-exponential coefficient = 0.3000
sum-of-squared residuals = 10066995
root mean squared error = 236.49

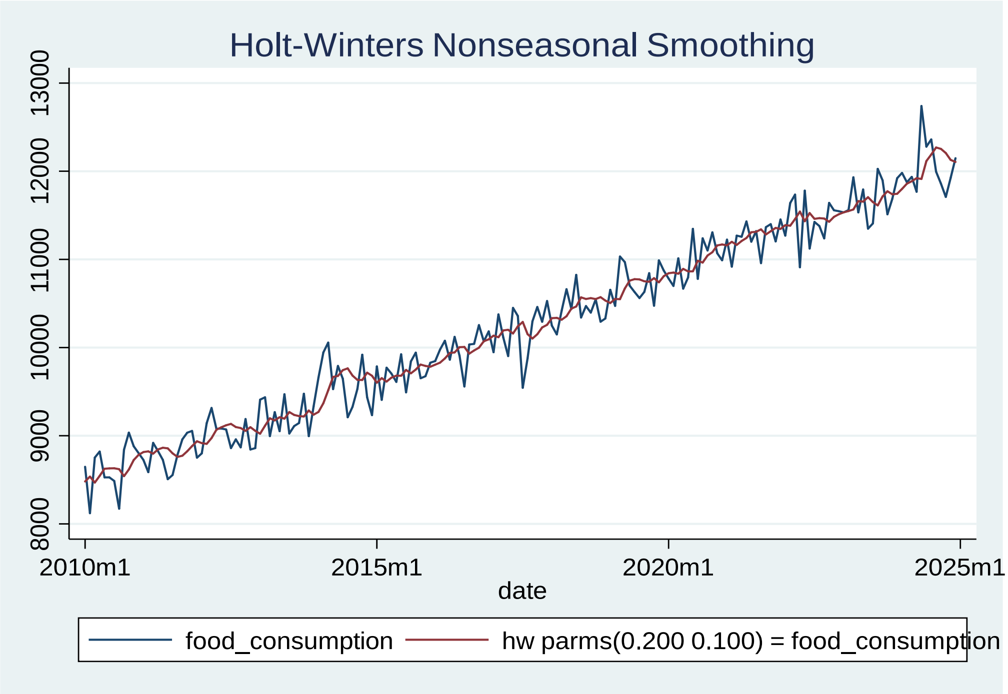


A better fit for trend, as it corrects for lag in the single-exponential model.

3. Holt-Winters Non-seasonal Smoothing

Specified weights:

alpha = 0.2000
beta = 0.1000
sum-of-squared residuals = 1.09e+07
root mean squared error = 245.557



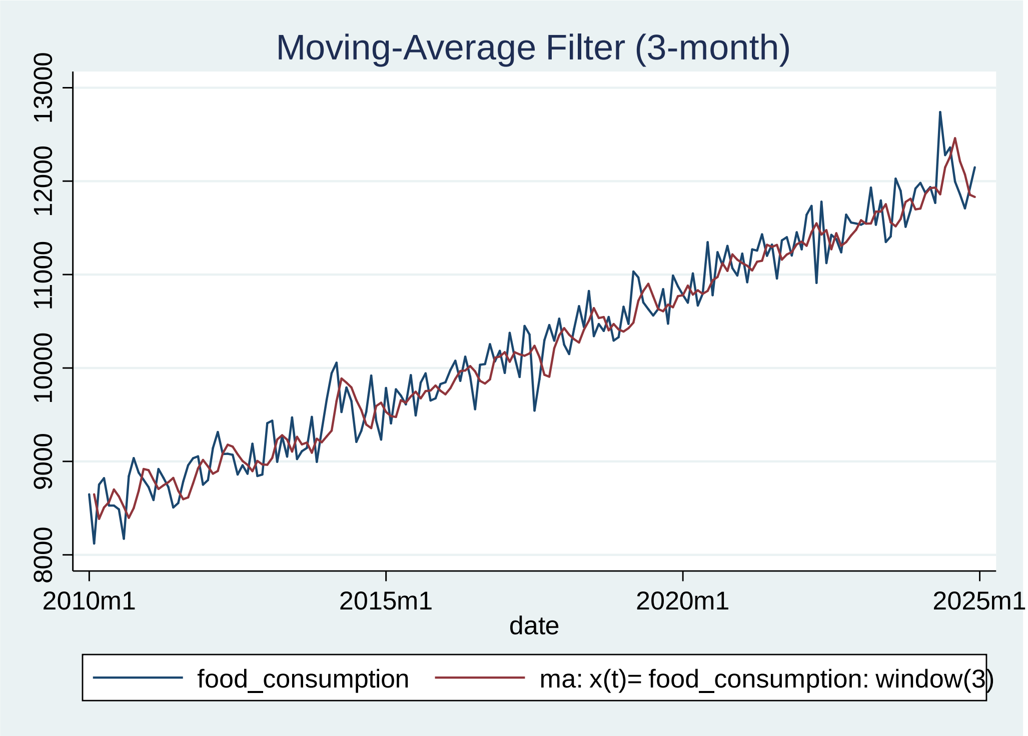
A **stronger trend component**, but **seasonality is not explicitly handled**.

4. Holt-Winters Seasonal Smoothing (Commented Out)

If seasonality is present, we would use this model that captures **both trend and seasonality**.

5 Moving-Average Filter (3-Month Window)

The smoother applied was
(1/3)*[x(t-3) + x(t-2) + x(t-1) + 0*x(t)]; x(t)= food_consumption



Here; removes short-term noise while preserving the overall trend.

Model Evaluation:

We calculate Mean Absolute Error (MAE) & Root Mean Squared Error (RMSE) for each method. The lowest MAE & RMSE indicate the best model for forecasting.

Model Evaluation:
Single-Exponential Smoothing - MAE: 210.22572, RMSE: 266.00336
Double-Exponential Smoothing - MAE: 189.11923, RMSE: 236.49048
Holt-Winters Non-seasonal - MAE: 199.8369, RMSE: 245.55701
Moving-Average (3-month) - MAE: 218.78062, RMSE: 266.17465

Forecasting for Next Year (2025)

- Extend Time-Series to 2025

We append 12 months to the dataset.

- Generate Forecast Using Best Model

A forecasted trend for 2025, where predictions continue the previous trend while adjusting based on the model.

STATA code: Deterministic Model

```
* Set time-series format
tsset date, month

* Summary statistics
summarize

* Plot original food consumption data
tsline food_consumption, title("Original Food Consumption Data")

* ---------------- Smoothing Techniques ---------------- *

* 1.Single-Exponential Smoothing
tssmooth exponential food_smooth_exp = food_consumption
tsline food_consumption food_smooth_exp, title("Single-Exponential Smoothing")

* 2. Double-Exponential Smoothing (Holt's Method)
tssmooth dexponential food_smooth_double = food_consumption, parm(0.3)
tsline food_consumption food_smooth_double, title("Double-Exponential Smoothing")

* 3. Holt-Winters Nonseasonal Smoothing
tssmooth hwinters food_smooth_hw = food_consumption, parms(0.2,0.1)
tsline food_consumption food_smooth_hw, title("Holt-Winters Nonseasonal Smoothing")

* 4. Holt-Winters Seasonal Smoothing
*tssmooth hwinters food_smooth_hw_season = food_consumption, parms(0.2, 0.1)
period(12)
*tsline food_consumption food_smooth_hw_season, title("Holt-Winters Seasonal
Smoothing")

* 5. Moving-Average Filter (3-month window)
tssmooth ma food_smooth_ma = food_consumption, window(3)
tsline food_consumption food_smooth_ma, title("Moving-Average Filter (3-month)")

* ---------------- Model Evaluation ---------------- *
* Compute residuals for each smoothing method
gen error_exp = food_consumption - food_smooth_exp
* Compute Mean Absolute Error (MAE)
egen mae_exp = mean(abs(error_exp))
* Compute Root Mean Squared Error (RMSE)
gen error_sq_exp = error_exp^2
egen rmse_exp = mean(error_sq_exp)
replace rmse_exp = sqrt(rmse_exp)

gen error_double = food_consumption - food_smooth_double
* Compute Mean Absolute Error (MAE)
egen mae_double = mean(abs(error_double))
* Compute Root Mean Squared Error (RMSE)
gen error_sq_double = error_double^2
egen rmse_double = mean(error_sq_double)
replace rmse_double = sqrt(rmse_double)
```

```
gen error_hw = food_consumption - food_smooth_hw
* Compute Mean Absolute Error (MAE)
egen mae_hw = mean(abs(error_hw))
* Compute Root Mean Squared Error (RMSE)
gen error_sq_hw = error_hw^2
egen rmse_hw = mean(error_sq_hw)
replace rmse_hw = sqrt(rmse_hw)

*gen error_hw_season = food_consumption - food_smooth_hw_season
* Compute Mean Absolute Error (MAE)
*egen mae_hw_season = mean(abs(error_hw_season))
* Compute Root Mean Squared Error (RMSE)
*gen error_sq_hw_season = error_hw_season^2
*egen rmse_hw_season = mean(error_sq_hw_season)
*replace rmse_hw_season = sqrt(rmse_hw_season)

gen error_ma = food_consumption - food_smooth_ma
* Compute Mean Absolute Error (MAE)
egen mae_ma = mean(abs(error_ma))
* Compute Root Mean Squared Error (RMSE)
gen error_sq_ma = error_ma^2
egen rmse_ma = mean(error_sq_ma)
replace rmse_ma = sqrt(rmse_ma)


* Display results
display "Model Evaluation:"
display "Single-Exponential Smoothing - MAE: " mae_exp ", RMSE: " rmse_exp
display "Double-Exponential Smoothing - MAE: " mae_double ", RMSE: " rmse_double
display "Holt-Winters Nonseasonal - MAE: " mae_hw ", RMSE: " rmse_hw
*display "Holt-Winters Seasonal - MAE: " mae_hw_season ", RMSE: " rmse_hw_season
display "Moving-Average (3-month) - MAE: " mae_ma ", RMSE: " rmse_ma

* ---------------- Forecasting for next year ---------------- *
tsappend, add(12)   // Add 12 more months for forecasting (2025)
* Generate forecast variable for next year
predict food_forecast, dynamic(180)
```

🕮 Regression with ARIMA Errors

Regression analysis is crucial for forecasting agricultural data as it helps identify relationships between variables, enabling accurate predictions of crop yields, pricing, and resource management. It enhances decision-making by modeling trends and forecasting future outcomes based on historical patterns.

Regression with ARIMA errors combines the strengths of regression analysis and time series modeling by addressing autocorrelation in residuals. This approach accounts for both the deterministic relationships between variables and the stochastic patterns in the error term, improving forecast accuracy. It is especially useful when residuals exhibit significant serial correlation, which standard regression models fail to capture.

🕮 Overview of ARIMA and Regression Models

a. Regression Models

Regression analysis is used to model the relationship between a dependent variable $Y_t$and one or more independent variables X_t. The general form of a linear regression model is:

$$Y_t = \beta_0 + \beta_1 X_t + \epsilon_t$$

Where,$Y_t$is the dependent variable at time t,X_t is the independent variable at time t, β_0 is the intercept, and$\beta_1$is the coefficient of the independent variable, ϵ_t is the error term. This model assumes that the error terms are independent and identically distributed (i.i.d). However, in time series data, these errors often exhibit autocorrelation, meaning the error at one time point is correlated with the error at previous time points.

b. ARIMA Model

ARIMA is a time series forecasting technique that models both the autoregressive (AR) and moving average (MA) components of a time series. The general form of an ARIMA model is denoted as ARIMA(p, d, q), where:

p: Number of autoregressive terms (AR),

d: Degree of differencing to make the series stationary,

q: Number of moving average terms (MA).

ARIMA models are powerful for forecasting data with temporal dependence, but they focus only on the autocorrelation of the series itself. In the case of **regression with ARIMA errors**, we introduce ARIMA modeling for the residuals (errors) from a regression model, providing a more accurate forecast that accounts for both the relationships between variables and the temporal structure of the errors.

📖 Steps to Implement Regression with ARIMA Errors for Forecasting

The regression with ARIMA errors approach involves fitting a standard regression model to time series data and modeling the residuals (errors) using an ARIMA process.

Step 1: Prepare the Data

Step 2: Fit the Regression Model

Step 3: Analyze the Residuals

Step 4: Fit the ARIMA Model to the Residuals

Step 5: Combine the Models and Forecast

Step 6: Evaluate Model Performance

Advantages of Regression with ARIMA Errors

a) **Improved Forecast Accuracy**: By modelling the autocorrelation in the residuals, this method provides more accurate forecasts compared to traditional regression models that ignore serial dependence in the errors.
b) **Flexibility**: It can handle both the relationship between dependent and independent variables as well as the temporal structure in the data.
c) **Capturing Hidden Patterns**: ARIMA modelling of residuals helps uncover hidden autocorrelation patterns that are not captured by the initial regression model.

Conclusion

Regression with ARIMA errors is a powerful technique for forecasting time series data that combines the strengths of both regression and ARIMA models. By modelling the autocorrelation structure in the residuals of the regression, this approach improves forecasting accuracy and ensures that temporal dependencies are properly accounted for. This method is applicable in a wide range of domains, from economics to agriculture, where time series data often exhibit complex patterns and trends.

References

1. Box, G.E.P., Jenkins, G.M., &Reinsel, G.C. (2015). *Time Series Analysis: Forecasting and Control*. 5th edition.
2. Hyndman, R.J., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*. 2nd edition.
3. Greene, W.H. (2018). *Econometric Analysis*. 8th edition.

Hands-on exercise of regression with ARIMA error for forecasting agricultural production

Problem statement: Rice in Bangladesh is grown in three distinct seasons, namely *Boro* (January to June), *Aus* (April to August), and *Aman* (August to December). Data of total land area (in lac Hectare) and total production (in lac Metric tons), yield rate (Metric tons per Hectare) of *Boro rice* from financial year 1970-71 to 2014-15 were compiled from published Yearbook of Agricultural Statistics in Bangladesh, Bangladesh Bureau of Statistics (Excel file: production). Establish a regression with ARIMA error model for forecasting Boro production for next five years.

Practicing SARIMA modelling for forecasting agricultural responses using STATA software

Prof. Dr. Sheikh Mohammad Sayem

Seasonal autoregressive moving average (SARIMA) is one of the popular methods of forecasting time series data.

SARIMA (p, d, q) (P, D, Q)$_s$ model is termed a seasonal autoregressive integrated moving average model, where p is the order of autoregressive, d is the order of integration, q is the order of moving average, P is the order of seasonal autoregressive, D is the order of seasonal integration, Q is the order of seasonal moving average and s is the length of seasonal period.

Seasonality is defined as a pattern that repeats itself over a fixed interval of time.If the pattern is consistent one, the autocorrelation coefficient at a lag of 12 months will have a high positive value which may indicate the presence of seasonality.

The Box-Jenkins methodology can be applied to fit best seasonal autoregressive integrated moving average model for time series forecasting. The method consists of three phases.

Phase 1: Model Identification

Phase 2: Estimation and Testing Parameter of the Model

Phase 3: Forecasting

Algorithm 1: Fitting ARIMA Model

Step 1: *Start Stata → Import or Entry Data* *Statistics → Time series → Setup and Utilities → Declare datasets to be time series data*

In "tsset-Declare dataset to be time series data

- Select time variable and time frame

  Click OK

Step 2: *Statistics → Time series → ARIMA and ARMAX model → click*

- Select dependent variable.
- Select order of ARIMA (p,d,q) specification
- Select order of SARIMA (P,D,Q,S) specification

Click OK

Step 3: *Statistics → Post estimation → Specification diagnostic and goodness of fit analysis*

(Information Criteria-AIC and BIC)

Click Launch / Information criteria (ic)

Click OK

- Click Save and assign tick marks on predicted values, lower confidence limits and upper confidence limits

- From Option, enter forecast point
- Select Plots especially tick Residual Autocorrelation Function and Residual Partial Autocorrelation Function
- Click OK

Problem: Monthly wholesale price of potato (Tk. /100kg) between the years January 2000 to December 2014 are given as a Microsoft Excel File namely “Potato Price_SARIMA”. The question is “How can one forecast future price of potato by using time series modeling?”

Phase 1: Firstly need to develop a time series plot to check the pattern of the data

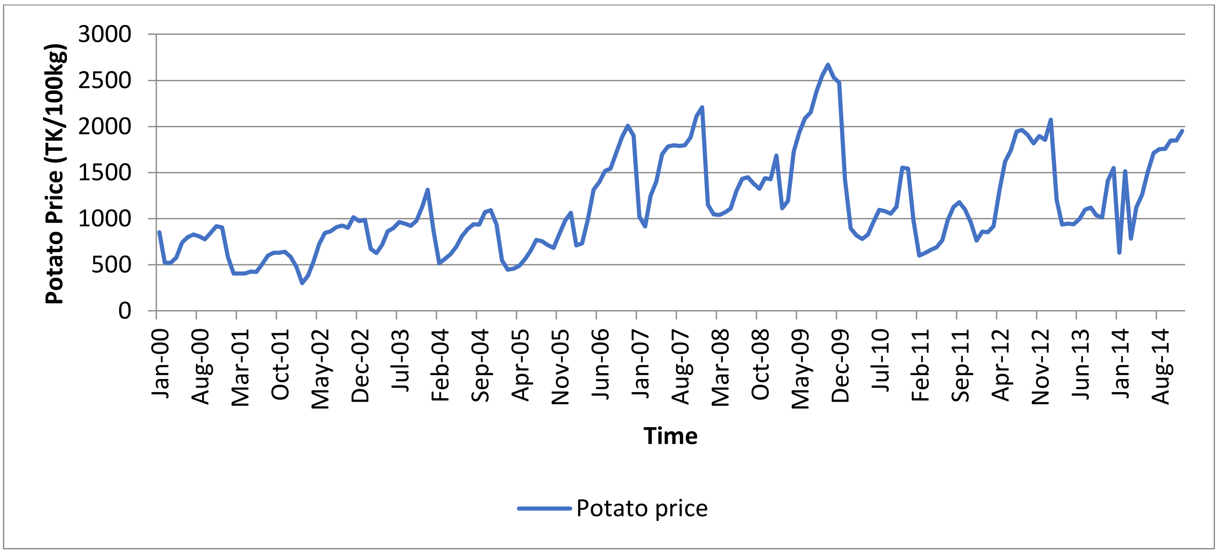
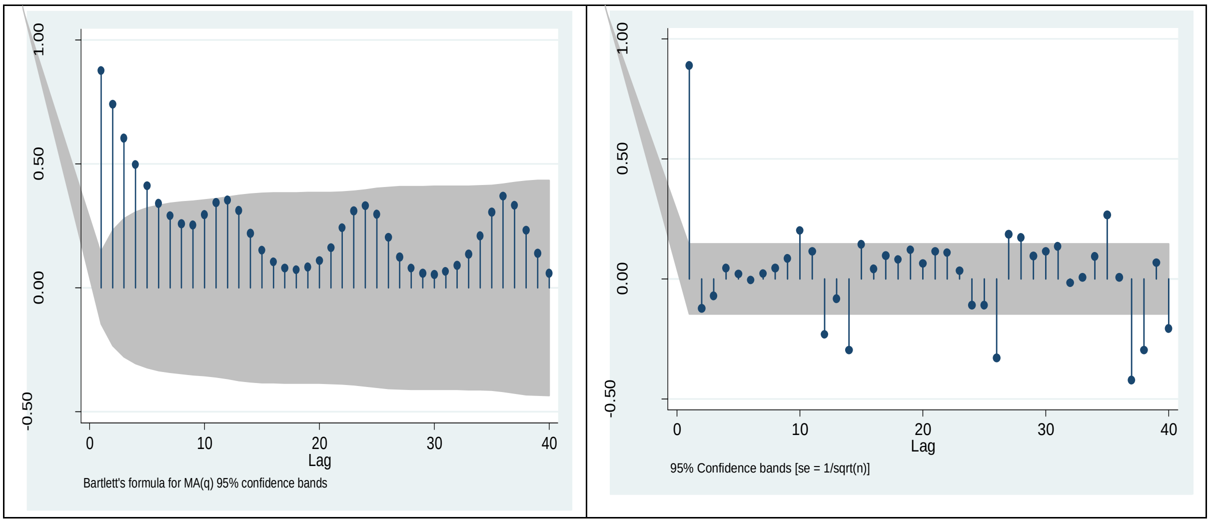


Figure: Time series plot of potato price from january, 2000 to December 2014

- The monthly potato price data of Bangladesh follows seasonality pattern (fig 1). Now, the task to define date of the data monthly. And check the stationarity by drawing the correlogram.



➢ The price data are not follow stationary time series data.

Table 2: Model selection criteria for checking the best fitted model for Potato price

Model	**AIC**	**BIC**	**Ljung-Box Q**	
			Statistic	P-value
SARIMA (1,0,0)(0,1,2)12	2394.58	2403.952	90.2883	0.0000
SARIMA (1,0,0)(0,1,1)12	**2296.526**	**2309.022**	43.4378	0.3305

Among these models, SARIMA (1, 0, 0) (0, 1, 1)12 has both lowest AIC and BIC values and appeared to be the best model. Moreover, the Ljung-Box test suggested that the ACF of residuals for the model at different lag times was not significantly different (no significant autocorrelation) from zero (Q=43.4378) and *p=0.3305>0.05)*.

```
ARIMA regression

Sample:  13 - 0180                              Number of obs     =        168
                                                Wald chi2(2)      =     586.24
Log likelihood = -1144.263                      Prob > chi2       =     0.0000
```

S12. Potatoprice	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
Potatoprice						
_cons	56.82742	21.8924	2.60	0.009	13.9191	99.73574
ARMA						
ar L1.	.8656017	.0357666	24.20	0.000	.7955006	.9357029
ARMA12						
ma L1.	-.9999737	325.322	-0.00	0.998	-638.6193	636.6193
/sigma	198.8255	32339.19	0.01	0.498	0	63582.48

The best selected SARIMA model for forecasting wholesale price of potato is SARIMA (1, 0, 0) (0, 1, 1)12 among the other tentatively selected SARIMA model based on the model selection criteria.

Basic concepts of dynamic regression models

Prof. Dr. Sheikh Mohammad Sayem

Dynamic Regression Modeling

A dynamic regression model is a regression model that allows lagged values of the explanatory variable(s) to be included. This model is used to predict what will happen to the forecast variable if the explanatory variable changes. Dynamic regression model is a mixed-model approach which can be applied to identify the role of explanatory variables in determining the variable of interest (dependent variable).

Steps of Dynamic Regression Model for Forecasting Time Series Data

Step 1: The first step in identifying the appropriate dynamic regression model is to fit a multiple regression model of the form-

$$Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_k X_{t-k} + N_t$$

Where Y_t refers to the forecasted variables;

X_{t-k} means the explanatory variable with time-lag *k=0, 1…K;*

$\beta_0, \beta_1, \ldots, \beta_k$ the regression coefficients and N_t is an ARIMA process.

Step 2: If the errors from the regression appear non-stationary, need to use differencing of the variable. Fit the model again using the lower-order ARIMA model for the errors.

Step 3: If the errors now appear stationary, identify the number of lagged explanatory variables that will be influenced forecast variable.

Step 4: Calculate the errors from the regression model and identify the appropriate ARMA model for the error series.

Step 5: Refit the entire model using the new ARMA model for the errors and the transfer function model for explanatory variables.

Step 6: Finally, check the adequacy of the fitted model by analyzing the residuals.

Forecasting Accuracy

In time series analysis, it is important to measure the suitability of a particular forecasting method for a given data set. Forecasting accuracy refers to how well the forecasting model can reproduce the data that are already known. Some important tools for measuring the forecasting accuracy are as follows:

Bayesian Information Criterion (BIC): The Bayesian Information Criterion (BIC) is a way of selecting a model from a set of models for a given set of data. It is commonly used with ARIMA models to determine the appropriate model order. The function of BIC is as follows-

$$BIC = -2\{Log(Likelihood)\} + d \times \log(N)$$

Where, N is the sample size and d is the total number of parameters.

Mean Absolute Percentage Error (MAPE): MAPE is one of the measures of accuracy that refers to the average of the sum of all of the percentage errors for a given data set taken without regard to sign. The function of MAPE is-

$$MAPE = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{\widehat{U}_t}{Y_t}\right| \times 100$$

Where, n is the number of observations, Y_t is the observed value and U_t is the difference between the observed and estimated values.

Root Mean Squared Error (RMSE): The MSE is a measure of accuracy computed by squaring the individual error for each item in a data set and then finding the average mean value of the sum of those squares.The RMSE is simply the square root of the MSE. The functional form of MSE is-

$$MSE = \frac{1}{n}\sum_{t=1}^{n}\widehat{U}_t^2$$

Where n refers the number of observations, U_t is the difference between the observed and estimated values.

The R^2 Criterion: The coefficient of determination (R^2) is a statistical model selection criteria whose functional form is as follows:

$$R^2 = 1 - \frac{\sum_{n=1}^{t}\widehat{U}_t^2}{\sum_{n=1}^{t}(Y_t - \bar{Y})^2}$$

Where, n is the number of observations, Y_t is the observed value of the dependent variable and U_t is the difference between the observed and estimated values.

Ljung-Box Test: The Ljung-Box testis a test for autocorrelated errors. The functional form of the test statistic is as follows-

$$Q = n(n+2)\sum_{k=1}^{h}(n-k)^{-1}\rho_k^2$$ Which follows χ^2 with the degree of freedom *h-m.*

Where n, k, h, m, and ρ_k refer to the number of observations, the number of lag, the maximum lag, the number of parameters and the autocorrelation function respectively. Ljung-Box test can be used to test the hypothesis that all of the autocorrelations are zero; that is, that the series is a white noise.

The correlogram view of the residuals (forecast errors) shows the autocorrelations of the residuals. Autocorrelation of the residual series falls outside the boundaries of the Correlogram suggesting there may be some additional information in the series which is not being captured by the forecasting model.

The best model can be found on the basis of maximum value of R^2, and minimum value of root mean squared error (RMSE), mean absolute percent error (MAPE) and Bayesian information criterion (BIC) (Gujarati, 2003).

Dynamic regression models using STATA software (ARDL Model)

Lagged Variables

- A possible source of any problem with the functional form is the lack of a lagged structure in the model.
- One way of overcoming autocorrelation is to add a lagged dependent variable to the model.
- However although lagged variables can produce a better functional form, we need theoretical reasons for including them.

Inclusion of Lagged variables

- Inertia of the dependent variable, whereby a change in an explanatory variable does not immediately effect the dependent variable.
- The overreaction to ‘news’, particularly common in asset markets and often referred to as ‘overshooting’, where the asset ‘overshoots’ its long-run equilibrium position, before moving back towards equilibrium
- To allow the model to produce dynamic forecasts.

Types of Lag

- Autoregressive refers to lags in the dependent variable

- Distributed lag refers to lags of the explanatory variables
- Moving average refers to lags in the error term

- Recent development in econometrics have however, revealed that often times, most time series are not stationary as was conventionally thought. Therefore, different time series may not display the same features. Hence, it is possible to see some time series that display the feature of diverging away from their mean over time while others may converge to their mean over time. Time series that diverge away from their mean over time are said to be non-stationary.
- The classical estimation of variables with this relationship most times gives misleading inferences or spurious regression.
- To overcome this problem of non-stationarity and prior restrictions on the lag structure of a model, econometric analysis of time series data has increasingly moved towards cointegration. The reason is that, cointegration is a powerful way of detecting the presence of steady state equilibrium between variables. Cointegration has become an over-riding requirement for any economic model using non-stationary time series data.
- If the variables do not cointegrate, then the problems of spurious regression and the results therein become almost meaningless. On the other hand, if the variables do cointegrate then we have cointegration.
- In applied econometrics, the Granger (1981) and, Engle and Granger (1987), Autoregressive Distributed Lag (ARDL) cointegration technique, Johansen and Juselius(1990) cointegration techniques have become the solution to determining the long run relationship between series that are non-stationary, as well as the Error Correction Model (ECM).
- The ECM result gives the short-run dynamics and long run relationship of the underlying variables.

ARDL Models

- An Autoregressive Distributed lag model or ARDL model refers to a model with lags of both the dependent and explanatory variables. An ARDL(1,1) model would have 1 lag on both variables

$$y_t = \alpha_0 + \alpha_1 x_t + \alpha_2 x_{t-1} + \alpha_3 y_{t-1} + u_t$$

- ARDL cointegration technique does not require pretests for unit roots unlike other techniques. Consequently, ARDL cointegration technique is preferable when dealing with variables that are integrated of different order, I(0), I(1) or combination of the both and, robust when there is a single long run relationship between the underlying variables in a small sample size.
- Although ARDL cointegration technique does not require pre-testing for unit roots, to avoid ARDL model crash in the presence of integrated stochastic trend of I(2), we are of the view the unit root test should be carried out to know the number of unit roots in the series under consideration.
- Since each of the underlying variables stands as a single equation, endogeneity is less of a problem in the ARDL technique because it is free ofresidual correlation (i.e. all variables are assumed exogenous). Also, it enable us analyze the reference model.
- The major advantage of ARDL lies in its identification of the cointegrating vectors where there are multiple cointegrating vectors.
- The Error Correction Model (ECM) can be derived from ARDL model through a simple linear transformation, which integrates short run adjustments with long run equilibrium without losing long run information.
- The associated ECM model takes a sufficient number of lags to capture the data generating process in general to specific modeling frameworks.

$$\Delta y_t = \alpha_0 + \alpha_1 t + \alpha_2 (y_{t-1} - \theta x_t) + \sum_{i=1}^{p-1} \psi_{yi} \Delta y_{t-i} + \sum_{i=0}^{q-1} \psi'_{xi} \Delta x_{t-i} + u_t$$

The steps of the ARDL Cointegration Approach

Step 1: Determination of the Existence of the Long Run Relationship of the Variables

At the first stage the existence of the long-run relation between the variables under investigation is tested by computing the Bound F-statistic (bound test for cointegration) in order to establish a long run relationship among the variables. This bound F-statistic is carried out on each of the variables as they stand as endogenous variable while others are assumed as exogenous variables.

In practice, testing the relationship between the forcing variable(s) in the ARDL model leads to hypothesis testing of the long-run relationship among the underlying variables.

The null of non-existence of the long-run relationship is defined by;

Ho: $\delta_1 = \delta_2 = 0$ (null, i.e. the long run relationship does not exist)

H1: $\delta_1 \neq \delta_2 \neq 0$ (Alternative, i.e. the long run relationship exists)

Step 2: Choosing the Appropriate Lag Length for the ARDL Model

If a long run relationship exists between the underlying variables, then ARDL approach to cointegration can be applied. The issue of finding the appropriate lag length for each of the underlying variables in the ARDL model is very important because we want to have standard normal error terms that do not suffer from non-normality, autocorrelation, heteroskedasticity etc.

- **Step 3: ARDL into Error Correction Model**
- As we discuss before, when non-stationary variables are regressed in a model we may get results that are spurious.
- One way of resolving this is to difference the data in order to achieve stationarity of the variables. In this case, the estimates of the parameters from the regression model may be correct and the spurious equation problem resolved.
- However, the regression equation only gives us the short-run relationship between the variables. It does not give any information about the long run behaviour of the parameters in the model. This create a problem since researchers are mainly interested in long-run relationships between the variables under consideration.
- To solve this problem the concept of cointegration and the ECM becomes imperative. With the specification of ECM, we now have both long-run and short-run information incorporated.

$$\Delta y_t = - \phi(1, \hat{p})EC_{t-1} + \sum_{i=1}^{k} \beta_{i0}\Delta x_{it} + \delta \Delta w_t - \sum_{j=1}^{p-1} \phi_j \Delta_{t-j} - \sum_{i=1}^{k} \sum_{j=1}^{\hat{q}-1} \beta_{ij}\Delta x_{i,1-j} + \mu_t$$

EC_t is the error correction term defined by;

$$EC_t = \mathcal{E}_t = y_t - \sum_{i=1}^{k} \hat{\theta}_i x_{it} - \psi' w_t$$

- The term EC_t as the speed of adjustment parameter or feedback effect is derived as the error term from the cointegration models. The EC_t shows how much of the disequilibrium is being corrected, that is, the extent to which any disequilibrium in the previous period is being adjusted in y_t.
- A positive coefficient indicates a divergence, while a negative coefficient indicates convergence. If the estimate of ECt = 1, then 100% of the adjustment takes place within the period, or the adjustment is instantaneous and full, if the estimate of $EC_t = 0.5$, then 50% of the adjustment takes place each period/year. $EC_t = 0$, shows that there is no adjustment, and to claim that there is a long-run relationship does not make sense any more.

- If the trace or Maximal eigen value or the F-statistics establishes that there exists a single long-run relation among the variables (i.e underlying variables), ARDL approach can be applied rather than applying Johansen and Juselius approach. The ARDL technique provides a unified framework for testing and estimating of cointegration relations in the context of a single equation
- When there are multiple long-run relationships, ARDL approach cannot be applied. Hence, an alternative approach like Johansen and Juselius (1990) becomes more appropriate.
- If one wants to understand the dynamic relationship between two variables, there is a number of possible cases:
- Both are I(0), i.e. stationary. Then an OLS on the variable levels will be unbiased and efficient.
- The variables are integrated of the same order (eg. I(1)) but not cointegrated.
- Appropriate differentiation (i.e. first difference for first order integration) allows for OLS estimation.
- The variables are integrated of the same order and co-integrated. Then a level OLS provides the long-run relationship, whereas an Error Correction Model (ECM) (which can be estimated using OLS) represents the short-run dynamics.

- Data might be of different orders and/or co-integrated (“things are not as clear cut”). ARDL analyzes both short-run dynamics and long-run relationships.

Auto regressive distributed lag model (ARDL) and its advantages

- Autoregressive Distributed Lag Models (ARDL) model plays a vital role when comes to a need to analyze an economic scenario. In an economy, change in any economic variables may bring change in another economic variable beyond time. This change in a variable is not reflected immediately, but it distributes over future periods. Not only macroeconomic variables, but other variables such as loss or profit earned by a firm in a year can also affect the brand image of an organization over the period.
- On the other hand, the ARDL model addresses the issue of collinearity by allowing the lag of the dependent variable in the model with other independent variables and their lags.
- Assumptions for ARDL Model

  The absence of autocorrelation is the very first requirement of ARDL. The model requires that the error terms should have no autocorrelation with each other.

  There should not occur any heteroscedasticity in the data. In simple terms, the variance and mean should remain constant throughout the model.

  The data should follow a normal distribution.

  Data should have stationary either on I(0) or I(1) or on both. In addition to this, if any of the variables in the data has stationary at I(2), ARDL Model cannot run.

References

Emeka Nkoro and Aham Kelvin Uko (2016) Autoregressive Distributed Lag (ARDL) cointegration technique: application and interpretation, Journal of Statistical and Econometric Methods, vol.5, no.4, 2016, 63-91.

Samuel Asumadu Sarkodie and Phebe Asantewaa Owusu (2020)How to apply the novel dynamic ARDL simulations (dynardl) and Kernel-based regularized least squares (krls) MethodsX 7, 101160.

Testing of co-integration and causality

Prof. Dr. Sheikh Mohammad Sayem

Granger Causality Test

The **Granger Causality Test** (introduced by Clive Granger in 1969) is a statistical hypothesis test used to determine whether one **time series can predict another**. **It does not imply true causality** but rather checks for **predictive relationships**.

A time series **X "Granger-causes" Y** if past values of **X help predict** future values of **Y**, beyond what past values of Y alone can do.

Dragonflies and Rainstorms: Before a rainstorm, dragonflies tend to fly lower than usual.This happens because air pressure drops before a storm, making it harder for dragonflies to stay at higher altitudes.

Causality vs. Predictability:

- Dragonflies do NOT cause rain (no true causality).
- However, their flight pattern can help predict rain.

Thus, low-flying dragonflies "Granger-cause" rainstormsmeaning that past observations of dragonflies flying low provide useful predictive information for an upcoming rainstorm.

Mathematical Interpretation

If we define:

D_t = Dragonfly flight altitude at time t

R_t = Rainstorm occurrence at time t

Then, using the Granger causality test, we check if past values of $D_t$significantly improve the prediction of D_t. If they do, then dragonfly flight patterns Granger cause rainstorms.

Mathematical Formulation

For thetwo time series $X_t$and Y_t, we estimate the following regressions:

(1) Predicting Y without X (Univariate Model)

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \epsilon_t$$

Where Y_t is dependent on its own past values

ϵ_t is the error term.

(2) Predicting Y with X (Bivariate Model)

$$Y_t = \alpha + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{j=1}^{q} \gamma_j X_{t-j} + \epsilon_t$$

If the coefficients γ_j are statistically significant, then **X Granger-causes Y**.

The same test can be applied in the opposite direction to check if **Y Granger-causes X**.

Interpreting Granger Causality Results

- ✓ If X Granger-causes Y, it suggests that X provides useful predictive information for Y.
- ✓ If Y Granger-causes X, it suggests that Y provides useful predictive information for X.
- ✓ If both are significant, there is bidirectional causality.
- ✓ If neither is significant, there is no causal relationship.

Limitations of the Granger Causality Test

- ⌦ Correlation does not imply true causation (e.g., missing variables could influence both X and Y).
- ⌦ Lag selection is crucialincorrect lag choice can lead to misleading results.
- ⌦ Requires stationary time series, meaning non-stationary data must be differenced first.

Check the Granger Causality

Scenario 1: Advertising and Sales Revenue

Scenario 2: Interest Rates and Inflation

Scenario 3: Stock Prices and Oil Prices

Introduction of error correction mechanism (ECM) in agricultural research

Prof. Dr. Sheikh Mohammad Sayem

Understanding Co-integration

Imagine a drunk man leaving a bar. His movement is completely random, making unpredictable steps forward, sideways, or even backward. This follows a random walk, where each step depends on the previous one, and the process is non-stationary.

Now, let's say he has a dog, which also follows a random walk. The dog sniffs around, moves unpredictably, and follows its own path. However, they are connected by a leash.

Unrestricted Random Walks:

- → Both the drunk man and his dog move randomly and independently.
- → Since they have no restrictions, their paths may diverge significantly.
- → This represents two non-stationary time series with no co-integration.

Restricted by the Leash:

- → In the park, dogs must be on a leash, meaning that while they both still move randomly, the distance between them cannot grow indefinitely.
- → The dog can move left or right, but it will always return close to the drunk man.
- → This represents co-integrationthe two variables (time series) are individually non-stationary, but their difference remains stationary.

Economic Example: Exchange Rate and Inflation

Exchange rates and inflation rates often follow non-stationary processes.

However, in the long run, economic theory suggests they should maintain an equilibrium relationship (e.g., Purchasing Power Parity).

If inflation rises in one country, the exchange rate should adjust correspondingly.

Co-integration

In time series analysis, many economic and financial variables exhibit non-stationarity, meaning their statistical properties change over time. However, even if two or more series are non-stationary, they might still have a stable long-term relationship. This phenomenon is called co-integration.Co-integration implies that although individual time series wander unpredictably, a certain **linear combination** of them is **stationary**. Consider two non-stationary time series, $Y_t$and X_t:

- If both $Y_t \sim I(1)$and $X_t \sim I(1)$(i.e., both have unit roots), their linear combination might be stationary.
- There exists a parameter θ, such that: $Z_t = Y_t - \theta X_t$where $Z_t \sim I(0)$(stationary).
- If such a θexists, then Y_t and $X_t$are co-integrated with co-integrating coefficient θ.

This means that while $Y_t$and $X_t$might drift apart in the short run, they will return to equilibrium in the long run.

If two-time series are individually integrated of order d but a linear combination of them is integrated of a lower order, then these time series are said to be co-integrated. That is, if (*X*,*Y*) are each integrated of order *d*, and there exist coefficients *a*,*b* such that *aX* + *bY* is integrated of order less than d, then *X* and *Y* are co-integrated.In practice, co-integration is often used for two I(1) series, but it is more generally applicable and can be used for variables integrated of higher order

Error Correction Model (ECM)

When two time series are co-integrated, the short-run deviations from the long-term relationship must be corrected over time. The Error Correction Model (ECM) captures both short-run and long-run dynamics:

$$\Delta Y_t = \alpha + \sum_{i=1}^{p} \beta_i \Delta Y_{t-i} + \sum_{j=1}^{q} \gamma_j \Delta X_{t-j} + \lambda\ (Y_{t-1} - \theta X_{t-1} + \epsilon_t$$

Where;

ΔY_t represents the short-term change inY_t.

ΔX_t represents the short-term change in X_t

$(Y_{t-1} - \theta X_{t-1})$ is the error correction term (ECT), representing the long-run deviation from equilibrium.

λ is the speed of adjustment; it determines how quickly the system corrects itself toward equilibrium. If λis negative and significant, it confirms the existence of co-integration and indicates how fast the dependent variable returns to its long-term equilibrium

Co-integration captures the long-term stability in relationships between variables.ECM explains short-term adjustments and ensures that deviations from equilibrium are corrected. Together, they provide a comprehensive framework for modeling economic and financial time series.

Testing for Co-integration

Several methods are used to test for co-integration:

I. Engle-Granger Two-Step Method

A two-step method proposed by Engle and Granger (1987):

Step 1: Estimate the Long-Run Relationship Using OLS

We estimate the equation:

$$Y_t = \alpha + \beta X_t + \epsilon_t$$

Where

Y_t and $X_t$are non-stationary variables

ϵ_t represents the residuals.

Step 2: Test the Residuals for Stationarity

If the residuals(ϵ_t) are stationary (I(0)), then $X_t$and Y_t are co-integrated.

Use the Augmented Dickey-Fuller (ADF) test to check stationarity.

Example: GDP and Consumption

GDP and Consumption are often non-stationary.If they are co-integrated, then despite short-term fluctuations, they maintain a long-term economic relationship.

Limitations

- ✓ Only works for two variables.
- ✓ Cannot handle multiple co-integrating relationships.

II. Johansen and Juselius Method

A multivariate approach introduced by Johansen and Juselius (1990).

How It Works:

- Uses a Vector AutoRegression (VAR) framework.
- Tests for multiple co-integrating relationships.
- Relies on two test statistics:
    1. Trace test: Tests for the number of co-integrating relationships.
    2. Maximum eigenvalue test: Determines the rank of the co-integration matrix.

Example: Money Supply and Inflation

- In macroeconomics, money supply and inflation are often co-integrated.
- The Johansen test can identify multiple relationships, making it useful in economic modeling.

Limitations

- ✓ Requires large sample sizes.
- ✓ Sensitive to the choice of lags.

III. Engle and Yoo Three-Step Method

An extension of the Engle-Granger method, adding a third step to improve accuracy.

Step 1: Estimate the Co-integration Relationship

Similar to Engle-Granger, first estimate the long-run relationship.

Step 2: Test for Co-integration

Check if the residuals are stationary.

Step 3: Error Correction Model (ECM)

Incorporates short-run dynamics by including an error correction term. Captures both long-run equilibrium and short-run fluctuations.

Example: Interest Rates and Stock Prices

- If interest rates increase, stock prices often adjust in the long run.
- The Engle-Yoo method accounts for both short-term deviations and long-run trends.

Limitations

- ✓ More computationally complex.
- ✓ Still assumes one co-integrating relationship.

In 2003, Granger and his collaborator Robert Engle were jointly awarded the Nobel Memorial Prize in Economic Sciences.

Conditions for Applying Co-integration Tests

Before applying co-integration tests, ensure:

1. The time series are non-stationary (they have unit roots).
2. A linear combination of the series is stationary (integrated of order 0, I(0)).
3. The relationship is economically meaningful (not a spurious correlation).

Limitations and Applications of Each Method

Method	Limitations	Applications
Engle-Granger	Only works for two variables	Stock prices and dividends, GDP and consumption
Johansen &Juselius	Needs large samples, complex interpretation	Money supply & inflation, macroeconomic modeling
Engle-Yoo 3-Step	More computationally intensive	Interest rates & stock prices, foreign exchange

Conclusions

- ✓ Co-integration helps in identifying long-term relationships in time series data.
- ✓ The drunk man and dog analogy illustrates how two random processes can be linked in the long run.
- ✓ Different testing methods (Engle-Granger, Johansen, Engle-Yoo) offer varied levels of accuracy and applicability.
- ✓ Co-integration is widely used in finance, economics, and econometrics to model equilibrium relationships.

Application of error correction mechanism and co-integration in agricultural research

Prof. Dr. Sheikh Mohammad Sayem

Problem Statement: This dataset simulates a classic cointegration scenario, modeling the relationship between a drunk man and his dog over 500 time periods. The drunk man's path follows a random walk, meaning each position is the sum of the previous position and a random shock. The dog's path, however, is influenced by the drunk man's previous position with some added noise, making it a cointegrated series. Both series are expected to be non-stationary, but the dog should maintain a long-run relationship with the drunk man, making this dataset ideal for cointegration tests (Engle-Granger, Johansen) and Vector Error Correction Models (VECM) to analyze their short- and long-term dynamics.

- Is the drunk man's path a non-stationary random walk?
- Is the dog's path also non-stationary, or does it follow a stationary process?
- Do the drunk man and dog move together in the long run, implying cointegration?
- Can we model the short-term and long-term dynamics between them using a VECM framework?

Step-1: Check the stationarity

ADF test of original variables

DrunkMan_Dog_cointegration_DTA.dta", clear

tsset time

dfullerdrunk_man // Augmented Dickey-Fuller test

dfuller dog

```
. dfuller drunk_man // Augmented Dickey-Fuller test

Dickey-Fuller test for unit root                   Number of obs   =       499

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.224            -3.440            -2.870            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.1975

.
. dfuller dog

Dickey-Fuller test for unit root                   Number of obs   =       499

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -2.399            -3.440            -2.870            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.1420
```

Original steps of drunk man and dog are non-stationary.

ADF after 1st difference

dfullerd.drunk_man
dfullerd.dog

```
. dfuller d.drunk_man

Dickey-Fuller test for unit root                   Number of obs   =       498

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -22.549            -3.440            -2.870            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

.
. dfuller d.dog

Dickey-Fuller test for unit root                   Number of obs   =       498

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)            -26.904            -3.440            -2.870            -2.570
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000
```

After 1st difference steps of drunk man and dog are stationary

Step-2: Test for Cointegration using Engle-Granger Method

EG test which uses residuals from an OLS regression:

regress dog drunk_man // Estimate long-run relationship

```
. // Step 2: Test for Cointegration using Engle-Granger Method
. // EG test which uses residuals from an OLS regression:
. regress dog drunk_man // Estimate long-run relationship
```

Source	SS	df	MS			
				Number of obs	=	500
				F(1, 498)	=	29067.85
Model	26855.644	1	26855.644	Prob > F	=	0.0000
Residual	460.099787	498	.923895154	R-squared	=	0.9832
				Adj R-squared	=	0.9831
Total	27315.7438	499	54.7409695	Root MSE	=	.96119

dog	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
drunk_man	.7934944	.0046541	170.49	0.000	.7843503	.8026385
_cons	.0835734	.106883	0.78	0.435	-.1264238	.2935706

predict residuals, resid // Extract residuals
dfuller residuals, lags(5) // If residuals are stationary, cointegration exists

```
. predict residuals, resid // Extract residuals

. dfuller residuals // If residuals are stationary, cointegration exists
```

Dickey-Fuller test for unit root Number of obs = 499

	Test Statistic	Interpolated Dickey-Fuller 1% Critical Value	5% Critical Value	10% Critical Value
Z(t)	-22.020	-3.440	-2.870	-2.570

MacKinnon approximate p-value for Z(t) = 0.0000

The residual is stationary. Therefore, co-integration exist.

Step 3: Error Correction Model (ECM)

```
gen d_dog = d.dog // First difference of dog
gen d_drunk_man = d.drunk_man // First difference of drunk man
gen ect = L.residuals // Error correction term (lagged residuals)
regress d_dogd_drunk_manect // ECM model with short-run adjustments
```

Source	SS	df	MS			
				Number of obs	=	499
				F(2, 496)	=	946.00
Model	465.38683	2	232.693415	Prob > F	=	0.0000
Residual	122.0038	496	.245975402	R-squared	=	0.7923
				Adj R-squared	=	0.7915
Total	587.390629	498	1.17949926	Root MSE	=	.49596

d_dog	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
d_drunk_man	.0038702	.0213005	0.18	0.856	-.0379801	.0457205
ect	-1.006518	.0231472	-43.48	0.000	-1.051997	-.9610398
_cons	.0421996	.0222312	1.90	0.058	-.0014792	.0858785

Intervention Analysis

Intervention analysis is a statistical technique used to examine the effects of a specific event or intervention on a time series. This technique is particularly useful when you want to understand how an external factor (e.g., policy change, promotional campaign, or economic shock) influences the behavior of a time series.

The goal is to determine whether the intervention leads to a significant change in the pattern of the series and, if so, quantify the effect.

In time series modeling, interventions are typically introduced by adding external variables (intervention variables) into the model. One common approach to model the intervention effect is by integrating it with **ARIMA** (AutoRegressive Integrated Moving Average) models, where the errors from the ARIMA model are used to assess the intervention impact.

Types of Interventions

a. Step Intervention (or Dummy Variable Intervention)

A step intervention occurs when an event or intervention causes an immediate change in the level of the time series, and this change persists over time. This can be modeled using a dummy variable that takes a value of 0 before the intervention and 1 after the intervention.

Mathematically, the intervention model can be expressed as:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \epsilon_t + \theta_t \epsilon_{t-1} + X_t \beta + y D_t$$

Where

y_t is the observed value at time t,

X_t is a vector of exogenous predictors (if any),

D_t is the dummy variable representing the intervention (1 for post-intervention, 0 for pre-intervention),

γ represents the effect of the intervention.

b. Pulse Intervention

A pulse interventionrepresents a sudden and temporary change in the time series that lasts for only one period. This type of intervention is often used when the impact of an event (e.g., a promotional event or strike) is expected to last for a very short time.

The model for a pulse intervention is similar to a step intervention but focuses on a temporary effect:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \epsilon_t + \theta_t \epsilon_{t-1} + X_t \beta + y P_t$$

Where$P_t$is the pulse variable, which is 1 during the intervention period and 0 otherwise.

c. Temporary Level Shift

A temporary level shift intervention causes a short-term change in the mean level of the series, after which the series returns to its previous level. This can be modelled by applying a shift to the level of the series over a defined period.

d. Trend Intervention

A trend intervention changes the underlying trend of the series over time. This is modelled by introducing a slope or a trend variable that alters the original time series trajectory.

📖 **ARIMA with Intervention**

When using ARIMA models to analyze a time series with an intervention, the general approach is to:

1. Fit an ARIMA model to the original time series to capture its natural patterns (e.g., autoregressive behavior, differencing, and moving averages).
2. Introduce the intervention by adding dummy variables or other predictors to capture the effect of the event or intervention on the series.
3. Estimate the model parameters, including the intervention effects. The coefficients of the intervention variables represent the change in the level, slope, or trend caused by the intervention.
4. After fitting the model, the residuals (errors) should be checked for randomness. If the residuals are still correlated, further model refinement may be required (such as adding more lags or different intervention variables).

📖 **Practical Steps in Intervention Analysis**

1. **Identify the intervention**: Clearly define the event or intervention, such as a policy change or economic shock. Determine the time of its occurrence and whether the intervention is expected to affect the level, trend, or seasonality of the time series.
2. **Model the time series**: Fit an ARIMA model to the time series to understand its underlying behavior (autocorrelation, differencing, etc.).
3. **Introduce the intervention**: Add relevant intervention variables (e.g., dummy, pulse, or trend variables) to the ARIMA model.
4. **Estimate the model parameters**: Fit the augmented ARIMA model using statistical software, like R or Python, and interpret the coefficients.
5. **Evaluate the results**: Analyze the intervention's impact on the series by examining the significance of the intervention coefficients.
6. **Forecast and validate**: Use the fitted model for forecasting, and assess the model's forecasting accuracy by comparing predictions with actual values.

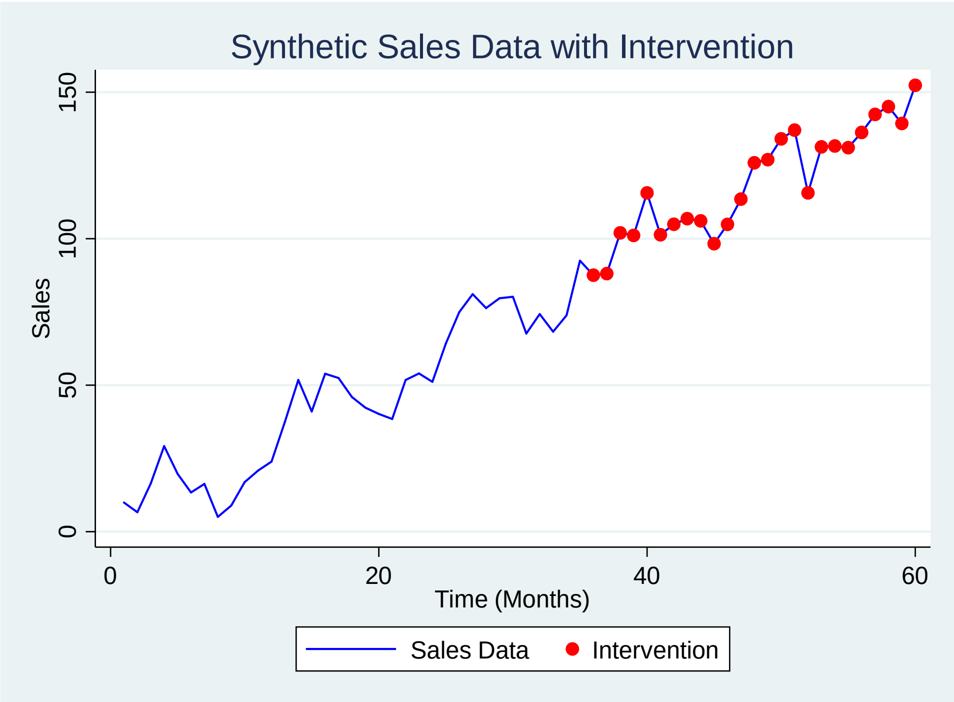
Conclusion

Intervention analysis in time series modelling is a powerful tool to assess the impact of external events or changes on a time series. By combining interventions with ARIMA models, we can more effectively quantify and forecast the effects of interventions.

References

1. Box, G.E.P., Jenkins, G.M., &Reinsel, G.C. (2015). *Time Series Analysis: Forecasting and Control*. 5th edition.
2. Harvey, A.C. (1990). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Problem Statement: This dataset represents simulated sales data over 60 time periods (which could be months, weeks, or days). The key feature of the data is an intervention that occurs at time = 36, potentially impacting sales. This artificial dataset models an intervention (policy change, new marketing strategy, product launch, or economic event) impacts sales growth. Here, Intervention is used as dummy variable: 0 (before intervention, periods 1-35), 1 (after intervention, periods 36-60).



Step-by-Step Analysis in Stata

Step 1: Load the Data

Step 2: Declare the Time-Series Structure

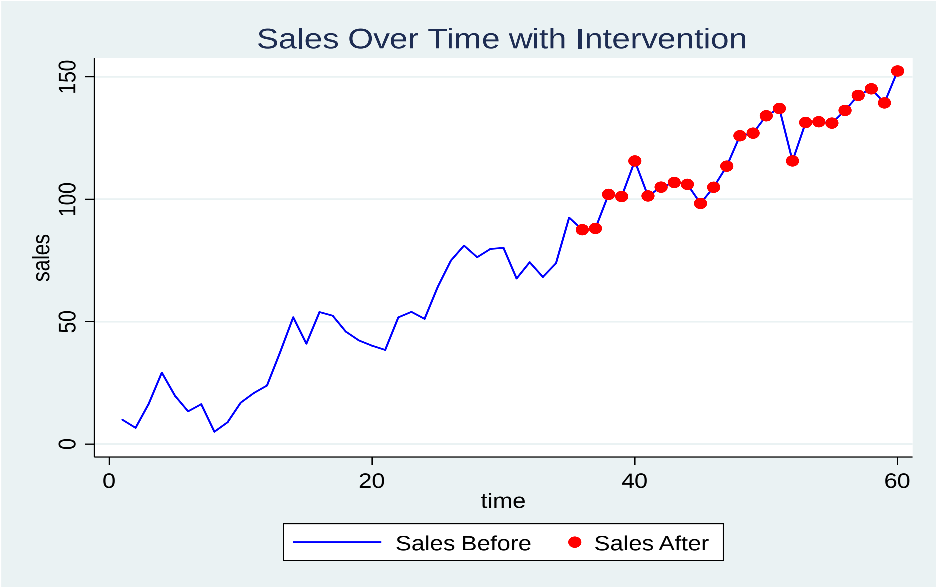
```
tsset time
```

This ensures Stata treats time as a **time-series variable**.

Step 3: Plot Sales Over Time

Visualizing the sales trend before and after intervention:

```
twoway (line sales time, sort lcolor(blue)) ///
    (scatter sales time if intervention == 1, mcolor(red)), ///
   title("Sales Over Time with Intervention") ///
   legend(label(1 "Sales Before") label(2 "Sales After"))
```



- A clear upward trend.
- A noticeable increase after the intervention (around time = 36).

Step 4: Compare Sales Before & After Intervention

Summary Statistics

summarize sales if intervention == 0 // Before intervention

summarize sales if intervention == 1 // After intervention

summarize sales if intervention == 0

Variable	Obs	Mean	Std. Dev.	Min	Max
sales	35	45.16726	25.67581	5.053794	92.50563

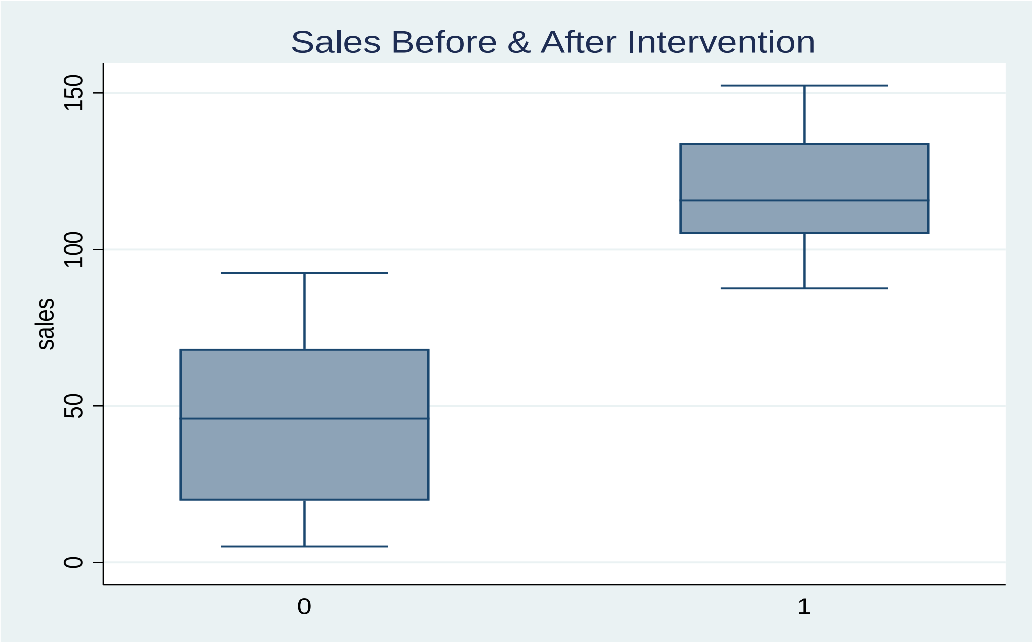
summarize sales if intervention == 1

Variable	Obs	Mean	Std. Dev.	Min	Max
sales	25	119.1784	18.52061	87.55426	152.3614

- sales before intervention: **lower mean, smaller range**.
- sales after intervention: **higher mean, larger range**.

Box Plot Comparison

```
graph box sales, over(intervention) title("Sales Before & After Intervention")
```



Two box plots showing **higher sales after the intervention**.

Step 5: T-Test (Did Sales Increase Significantly?)

ttest sales, by(intervention)

Two-sample t test with equal variances

Group	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
0	35	45.16726	4.340004	25.67581	36.34731	53.98721
1	25	119.1784	3.704122	18.52061	111.5335	126.8234
combined	60	76.00525	5.587749	43.28251	64.8242	87.18631
diff		-74.01118	6.01935		-86.06022	-61.96214

diff = mean(0) - mean(1) t = -12.2955

Ho: diff = 0 degrees of freedom = 58

Ha: diff < 0 Pr(T < t) = 0.0000

Ha: diff != 0 Pr(|T| > |t|) = 0.0000

Ha: diff > 0 Pr(T > t) = 1.0000

- If $p < 0.05$, the sales difference is **statistically significant**.

Step 6: Regression Analysis

To **quantify** the impact of the intervention:

reg sales intervention

```
. * Linear Regression: Impact of intervention
. regress sales time intervention
```

Source	SS	df	MS			
				Number of obs	=	60
				F(2, 57)	=	734.56
Model	106400.948	2	53200.4739	Prob > F	=	0.0000
Residual	4128.23875	57	72.4252412	R-squared	=	0.9627
				Adj R-squared	=	0.9613
Total	110529.187	59	1873.37604	Root MSE	=	8.5103

sales	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
time	2.333511	.1219496	19.14	0.000	2.089311	2.577711
intervention	4.005859	4.283788	0.94	0.354	-4.572279	12.584
_cons	3.164068	2.624448	1.21	0.233	-2.091298	8.419435

- **Intercept:** Baseline sales level.
- **Intervention Coefficient:** The average increase in sales after the intervention.
- **p-value:** If < 0.05, intervention **significantly impacted sales**.

Step 7: Time-Series Model

Stationarity: At original data

* Set time series structure
tsset time
* Check for stationarity using the Augmented Dickey-Fuller test
dfuller sales, regress lags(0)

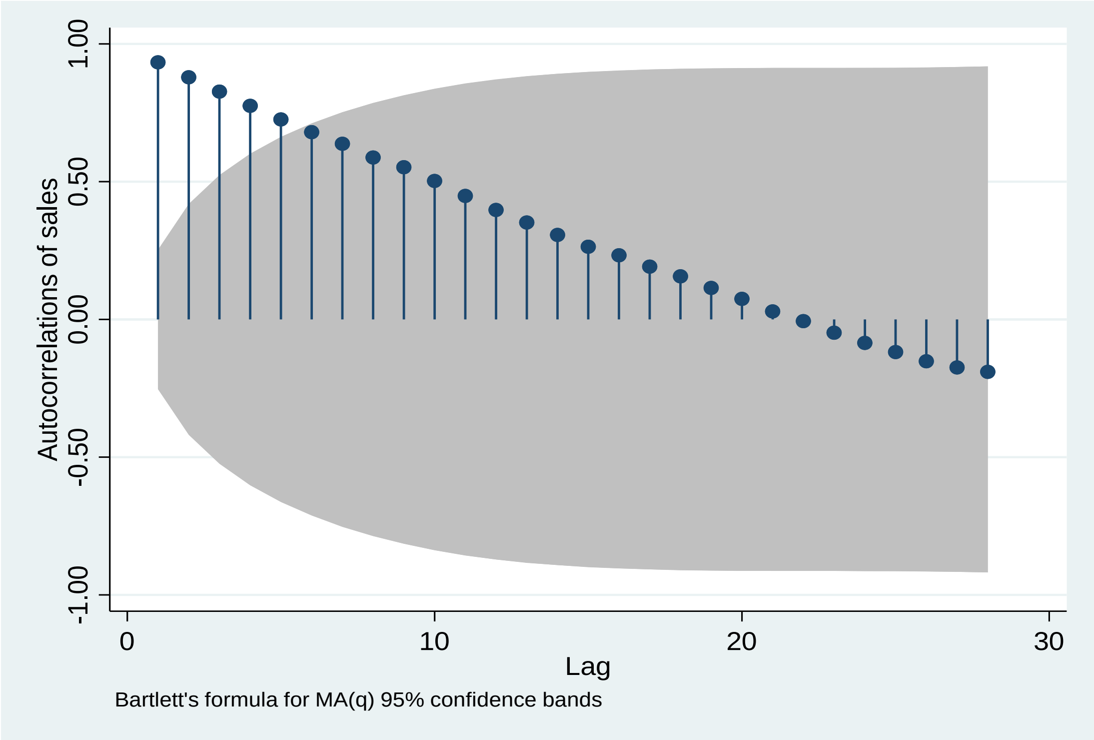
```
. dfuller sales, regress lags(0)

Dickey-Fuller test for unit root                   Number of obs   =        59

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -0.491            -3.567            -2.923            -2.596
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.8937

------------------------------------------------------------------------------
     D.sales |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       sales |
         L1. |  -.0130362   .0265467    -0.49   0.625     -.066195    .0401227
             |
       _cons |   3.387633   2.276622     1.49   0.142    -1.171225    7.946491
------------------------------------------------------------------------------
```

At first difference

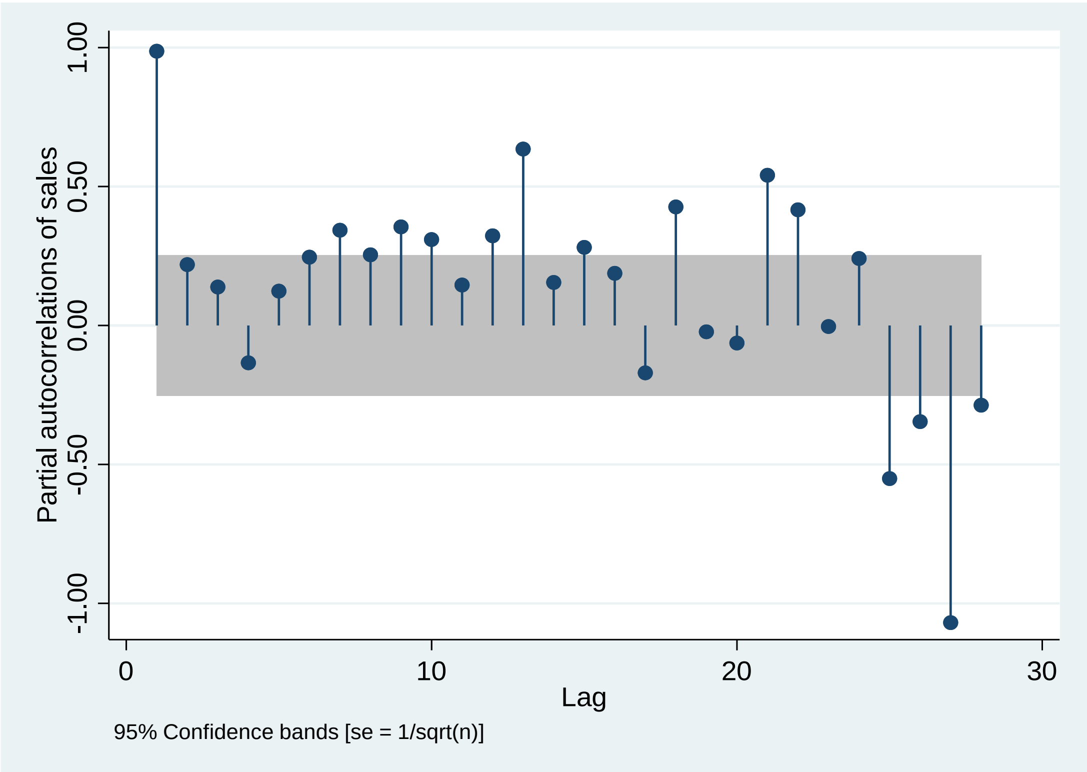
* If non-stationary, difference the series and test again
gen d_sales = d.sales
dfullerd_sales, regress lags(0)

```
. dfuller d_sales, regress lags(0)

Dickey-Fuller test for unit root                   Number of obs   =        58

                               ---------- Interpolated Dickey-Fuller ---------
                  Test         1% Critical       5% Critical      10% Critical
               Statistic           Value             Value             Value
------------------------------------------------------------------------------
 Z(t)             -9.317            -3.569            -2.924            -2.597
------------------------------------------------------------------------------
MacKinnon approximate p-value for Z(t) = 0.0000

------------------------------------------------------------------------------
   D.d_sales |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     d_sales |
         L1. |  -1.225574   .1315386    -9.32   0.000    -1.489077   -.9620701
             |
       _cons |   3.015678   1.144813     2.63   0.011     .7223421    5.309014
------------------------------------------------------------------------------
```

Autocorrelation Function:
ac sales

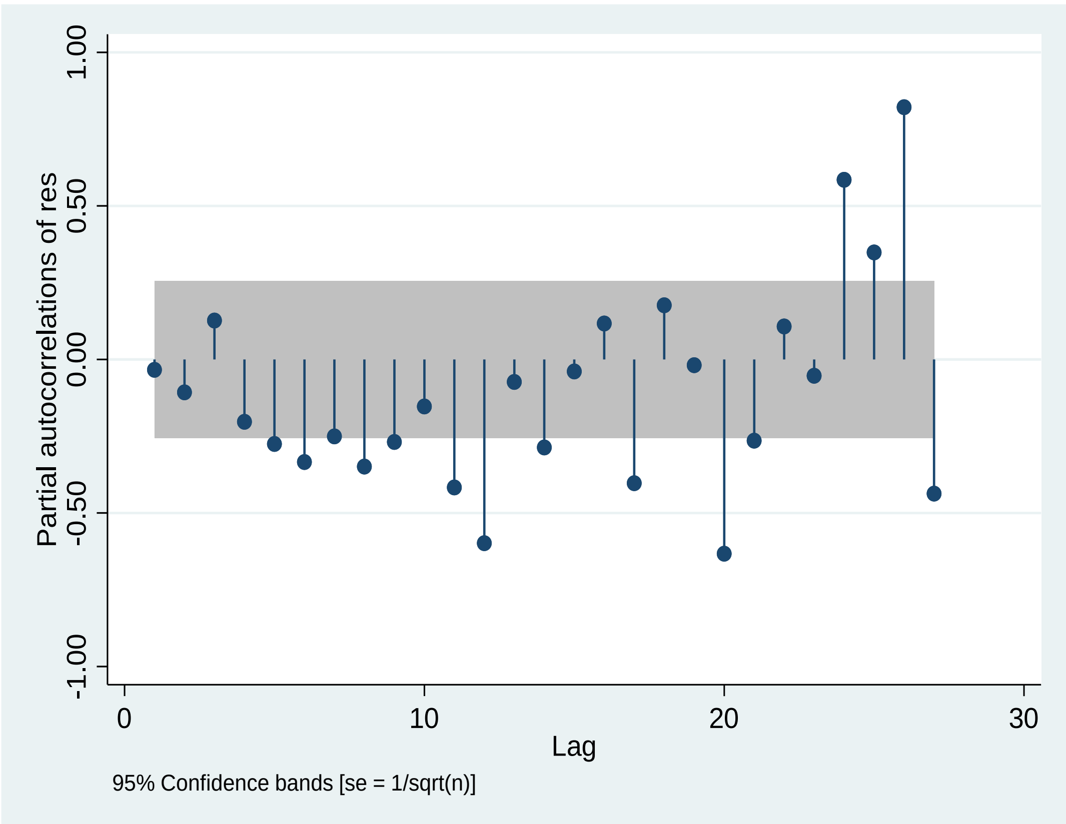


Partial ACF (PACF) function:
pac sales



* Step 3: Fit the best ARIMA model
arimad_sales, ar(1)

```
ARIMA regression

Sample:  2 - 60                                 Number of obs     =         59
                                                Wald chi2(1)      =       2.21
Log likelihood = -208.1937                      Prob > chi2       =     0.1372

------------------------------------------------------------------------------
             |                 OPG
     d_sales |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
d_sales      |
       _cons |   2.397151   .9079806     2.64   0.008     .6175419     4.17676
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |  -.2233502    .150258    -1.49   0.137    -.5178504    .0711501
-------------+----------------------------------------------------------------
      /sigma |   8.241742   .8022754    10.27   0.000     6.669311    9.814173
------------------------------------------------------------------------------
Note: The test of the variance against zero is one sided, and the two-sided
      confidence interval is truncated at zero.
```



* Check residual diagnostics
predict res, residuals
ac res
pac res

8. **ARIMA Model with intervention**
* ARIMA model for trend & intervention effect
tsset time
arimad_sales time intervention, ar(1)

```
ARIMA regression

Sample:  2 - 60                                 Number of obs     =         59
                                                Wald chi2(3)      =       2.43
Log likelihood = -208.1129                      Prob > chi2       =     0.4883
```

d_sales	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
d_sales						
time	.0407726	.0992297	0.41	0.681	-.153714	.2352593
intervention	-1.104724	3.40852	-0.32	0.746	-7.785301	5.575853
_cons	1.602294	2.224112	0.72	0.471	-2.756885	5.961474
ARMA						
ar L1.	-.2195084	.1540524	-1.42	0.154	-.5214455	.0824287
/sigma	8.231628	.8647863	9.52	0.000	6.536678	9.926578

Note: The test of the variance against zero is one sided, and the two-sided confidence interval is truncated at zero.

STATA code: Intervention Analysis

```
* Summary statistics
summarize sales if intervention == 0
summarize sales if intervention == 1

* Time-series visualization
twoway (line sales time, sort lcolor(blue)) ///
      (scatter sales time if intervention == 1, mcolor(red)), ///
      title("Sales Over Time with Intervention") ///
      legend(label(1 "Sales Before") label(2 "Sales After"))

*Box Plot Comparison
graph box sales, over(intervention) title("Sales Before & After Intervention")

* Check mean difference before & after intervention

ttest sales, by(intervention)

* Linear Regression: Impact of intervention
regress sales time intervention

* Identify the order of ARIMA
* Set time series structure
tsset time
* Check for stationarity using the Augmented Dickey-Fuller test
dfuller sales, regress lags(0)

* If non-stationary, difference the series and test again
gen d_sales = d.sales
dfullerd_sales, regress lags(0)

* ACF & PACF
ac sales
pac sales

* Step 3: Fit the best ARIMA model
arimad_sales, ar(1)

* Step 4: Check residual diagnostics
predict res, residuals
ac res
pac res

* ARIMA model for trend & intervention effect
tsset time
arimad_sales time intervention, ar(1)

* Forecast next 12 periods
tsappend, add(12)
predict sales_forecast, dynamic(60)
tsline sales sales_forecast, title("Forecasted Sales After Intervention")

display "Analysis Complete!"
```

Key concepts and applications of the ARCH model for forecasting agricultural data

Prof. Dr. Sheikh Mohammad Sayem

Agricultural data, such as crop yields, commodity prices, and weather patterns, often exhibit volatility and clustering effects. Traditional time series models, such as ARMA (Autoregressive Moving Average) and ARIMA (Autoregressive Integrated Moving Average), assume constant variance, which may not be suitable for highly volatile data. To address this, the **Autoregressive Conditional Heteroscedasticity (ARCH) model** and its extension, the **Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model**, are commonly used to capture time-varying variance.

🕮 Concepts of the ARCH Model

The ARCH model, introduced by Robert Engle (1982), captures time-dependent volatility by modelling the variance of a time series as a function of past squared errors. A basic **ARCH(q)** model is defined as:

$$X_t = \mu + \epsilon_t, \qquad \epsilon_t = \sigma_t Z_t, \qquad Z_t \sim N(0,1)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 \; ; \; \alpha_0 > 0, \alpha_i \geq 0$$

Where

X_t is the observed value,

ϵ_t is the error term,

σ_t^2 is the conditional variance,

Z_t is a white noise process,

α_i are the ARCH parameters.

🕮 Applications in Agricultural Forecasting

a) Commodity Price Volatility: ARCH models help in analyzing fluctuating agricultural commodity prices, such as wheat, rice, and soybeans.

b) Weather Impact on Crop Yields: By modelling variance, ARCH aids in assessing uncertainty in yield forecasts due to unpredictable weather.

c) Risk Management in Agricultural Investments: Understanding volatility patterns supports decision-making in agribusiness and insurance sectors.

🕮 Limitations of the ARCH Model

- ✓ Requires a large number of parameters for high-order models.
- ✓ Often inadequate for long-memory volatility processes.
- ✓ May produce negative variance estimates if parameters are not properly constrained.

Introduction of GARCH model for forecasting agricultural data

🕮 Concept of the GARCH Model

To overcome the limitations of ARCH, Bollerslev (1986) introduced the GARCH (Generalized ARCH) model, which incorporates past conditional variances.

A **GARCH(p, q)** model is formulated as:

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

Where

β_j represents past conditional variances,

p and q are the GARCH and ARCH orders, respectively.

🕮 Advantages over ARCH

- ✓ More parsimonious with fewer parameters.
- ✓ Better captures persistence in volatility.
- ✓ Suitable for modeling financial and agricultural time series data.

🕮 Applications in Agricultural Forecasting

- ✓ Yield Forecasting with Climatic Variability: GARCH models provide better estimates of uncertainty in agricultural productivity.
- ✓ Price Fluctuation Analysis: Helps in forecasting future commodity price volatility, supporting policy-making and hedging strategies.
- ✓ Market Risk Assessment: Essential for designing optimal crop insurance policies and investment strategies in agribusiness.

Conclusion

The ARCH and GARCH models offer robust techniques for forecasting time-varying volatility in agricultural data. While ARCH models effectively capture short-term volatility, GARCH models are more efficient for long-memory volatility structures. These models significantly contribute to price forecasting, yield analysis, and risk management in agricultural sectors, improving decision-making processes for farmers, policymakers, and investors.

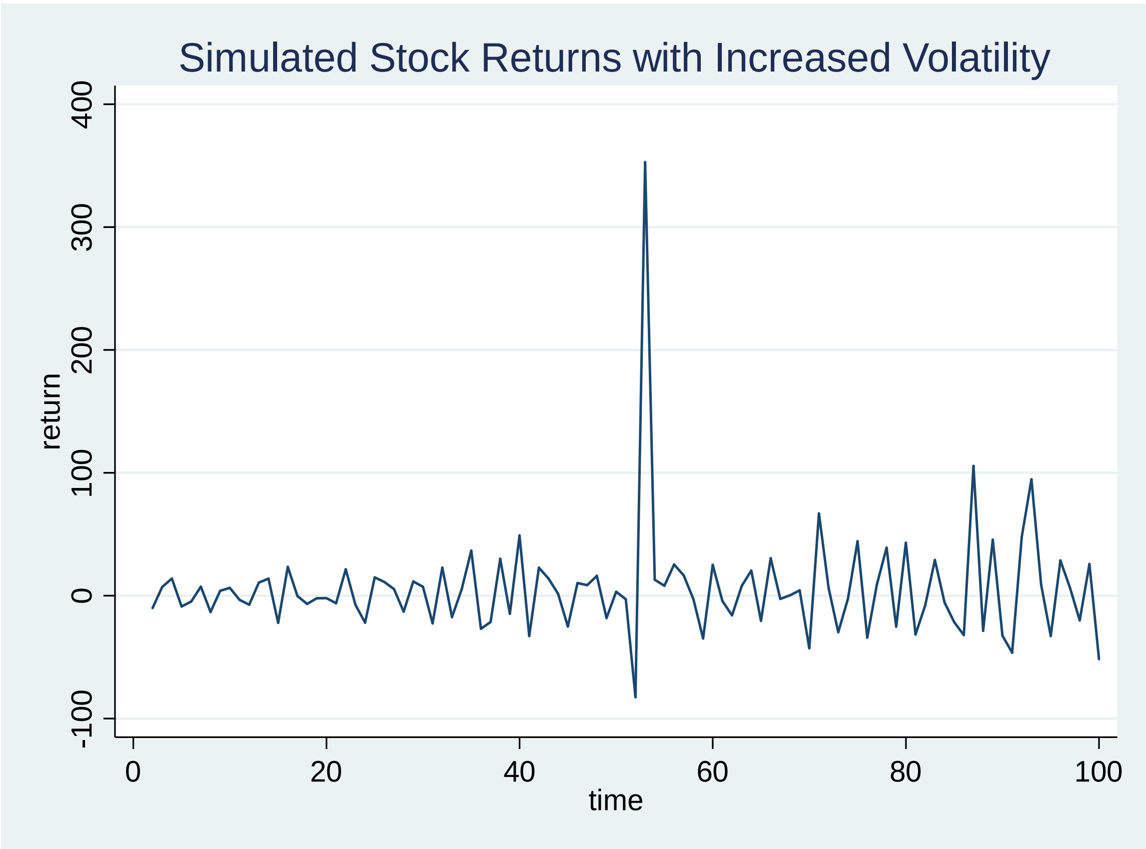
Hands-on exercise of ARCH and GARCH model for forecasting agricultural data using STATA software

Prof. Dr. Sheikh Mohammad Sayem

Problem statement: This dataset simulates stock price movements over 100 time periods (e.g., months) to analyze volatility dynamics in financial markets. The stock price series follows a random walk with increasing volatility over time, modeled using a volatility factor that grows linearly. Each stock price value is generated as a function of this factor and a normally distributed random shock. Stock returns are then computed as the percentage change from the previous period's price. This dataset is ideal for analyzing volatility clustering, performing ARCH/GARCH modeling, and testing for heteroscedasticity in financial time series.

Visualize the stock returns to check for volatility clustering

twoway (line return time), title("Simulated Stock Returns with Increased Volatility")



* Step 5: Check for heteroscedasticity in the returns (ARCH test)
arch return // ARCH test to check for volatility clustering

```
Time-series regression

Sample: 2 - 100                                   Number of obs   =        99
Distribution: Gaussian                            Wald chi2(.)    =         .
Log likelihood = -517.2256                        Prob > chi2     =         .
```

return	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
_cons	5.705541	7.34066	0.78	0.437	-8.681888	20.09297
/SIGMA2	2020.551	110.4305	18.30	0.000	1804.111	2236.991

*** Fit an ARCH(1) model**

arch return, arch(1)

```
ARCH family regression

Sample: 2 - 100                                   Number of obs   =        99
Distribution: Gaussian                            Wald chi2(.)    =         .
Log likelihood = -479.9859                        Prob > chi2     =         .
```

return	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
return						
_cons	3.180654	2.6702	1.19	0.234	-2.052842	8.41415
ARCH						
arch L1.	1.041735	.1712358	6.08	0.000	.7061192	1.377351
_cons	456.1914	63.15199	7.22	0.000	332.4158	579.967

*** Fit a GARCH (1, 1) model**

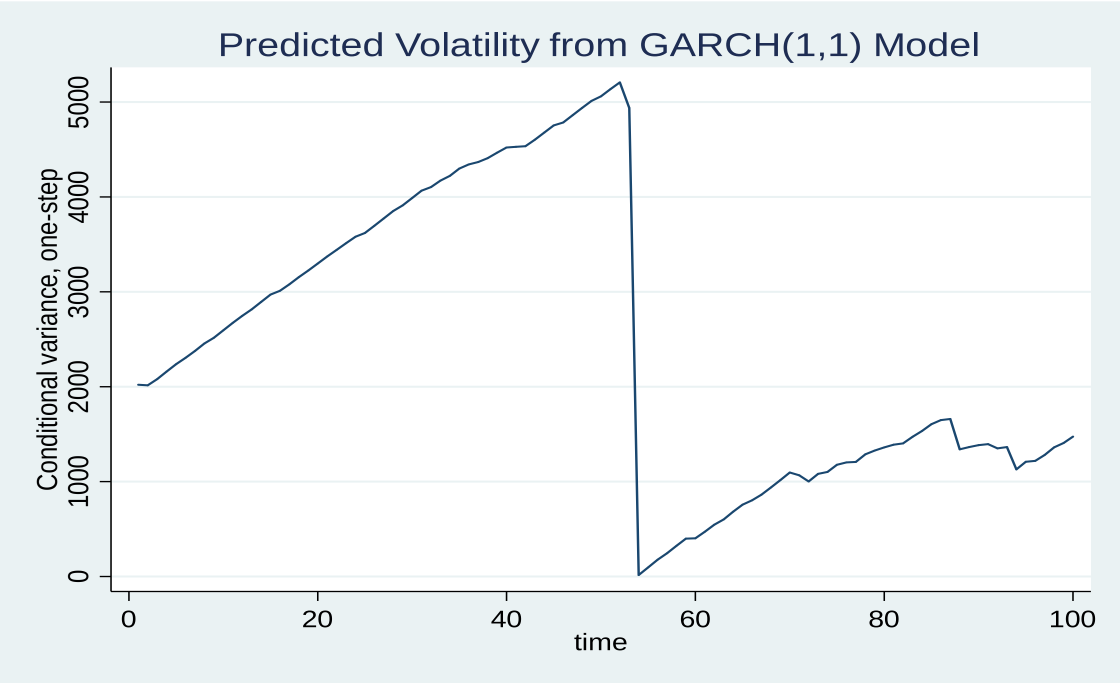
arch return, arch(1) garch(1)

```
ARCH family regression

Sample: 2 - 100                                 Number of obs     =         99
Distribution: Gaussian                          Wald chi2(.)      =          .
Log likelihood = -502.9733                      Prob > chi2       =          .

------------------------------------------------------------------------------
             |                 OPG
      return |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
return       |
       _cons |   8.132295   3.116173     2.61   0.009     2.024707    14.23988
-------------+----------------------------------------------------------------
ARCH         |
        arch |
         L1. |  -.0420523   .0038939   -10.80   0.000    -.0496842   -.0344205
             |
       garch |
         L1. |   .9991165   .0052226   191.31   0.000     .9888803    1.009353
             |
       _cons |   81.21249   17.34798     4.68   0.000     47.21107    115.2139
------------------------------------------------------------------------------
```

*** Step 9: Visualize the predicted volatility from the GARCH model**

predict volatility, variance

twoway (line volatility time), title("Predicted Volatility from GARCH(1,1) Model")



Key concepts of different non-linear econometric models in agricultural research

Prof. Dr. Sheikh Mohammad Sayem

A nonlinear econometric model refers to a statistical model where the relationship between the dependent and independent variables is not a straight line, meaning it cannot be expressed as a simple linear equation. These models are used to capture complex interactions, thresholds, and diminishing or increasing effects that are often present in real-world data.

Nonlinear models are essential when data shows patterns that linear models cannot adequately represent, such as saturation or exponential growth. They provide a more flexible approach to modelling phenomena like agricultural production, commodity pricing, and population growth. By allowing for more intricate relationships between variables, nonlinear models offer greater precision in forecasting and understanding dynamic systems.

🕮 Testing the nonlinearity

Testing the nonlinearity of an econometric model typically involves several approaches to check if a linear specification adequately captures the data's underlying structure:

1. **Residual Analysis**: The first step is to examine the residuals from a linear model. If patterns or systematic structures are present (such as trends or heteroscedasticity), it suggests nonlinearity.
2. **Ramsey RESET Test**: This test involves adding higher-order terms (squared or cubed predictors) to the linear model and checking for significant changes in model fit. A significant result indicates potential nonlinearity.
3. **Nonlinear Specification Tests**: Tests like the BDS test (Brock, Dechert, and Scheinkman) or the Harvey Collier test assess whether the residuals exhibit characteristics of nonlinearity or dependence.
4. **Likelihood Ratio Test**: Compare a linear model with a non-linear model (e.g., log-linear or polynomial models) to determine if the non-linear model significantly improves model fit.

5. **Graphical Methods**: Visualizing scatter plots or plotting the relationship between predictors and the dependent variable can provide intuitive evidence of nonlinearity.

📖 **Different Nonlinear econometric model in agricultural research**

1. Autoregressive Conditional Heteroscedasticity (ARCH) and Generalized ARCH (GARCH) Models

ARCH models capture time-dependent volatility, while GARCH models extend this by incorporating past variances. These models are widely used to analyze price volatility and risk in agriculture.

Applications in Agriculture:

- ✓ Forecasting price volatility in agricultural markets (e.g., wheat, rice, coffee).
- ✓ Assessing financial risks in agricultural investment and insurance.
- ✓ Modeling weather-driven yield variability.

2. Threshold Autoregressive (TAR) and Smooth Transition Autoregressive (STAR) Models

These models account for regime-switching effects. TAR models shift between regimes based on a threshold variable, while STAR models allow for smooth transitions.

Applications in Agriculture:

- ✓ Modeling price transmission asymmetries in food supply chains.
- ✓ Analyzing the effects of policy changes on agricultural production.
- ✓ Studying the impact of climate shocks on farm incomes.

3. Markov Switching Models (MSM)

These models assume that agricultural systems shift between hidden states (e.g., high and low volatility) based on a probabilistic rule.

Applications in Agriculture:

- ✓ Identifying structural breaks in commodity price cycles.
- ✓ Detecting regime shifts in farm productivity trends.
- ✓ Modeling transitions in agricultural trade policies.

4. Nonlinear Panel Data Models

Nonlinear panel models account for heterogeneity across farms, regions, or countries, using methods such as random-coefficient models, fixed-effects nonlinear models, and dynamic panel models**.**

Applications in Agriculture:

- ✓ Estimating farm productivity differences across regions.
- ✓ Analyzing the impact of government subsidies on farm efficiency.
- ✓ Understanding technological adoption in small vs. large-scale farms.

5. Structural Nonlinear Regression Models

These models explicitly define economic relationships based on agricultural theory, including Cobb-Douglas, Translog, and Quadratic production functions.

Applications in Agriculture:

- ✓ Estimating the effects of fertilizer, irrigation, and labor on crop yields.
- ✓ Analyzing supply and demand elasticity for agricultural commodities.
- ✓ Modeling farm income distribution under different market conditions.

6. Artificial Neural Networks (ANN) and Machine Learning Models

Machine learning models, such as Artificial Neural Networks (ANN), Support Vector Machines (SVM), and Decision Trees, are powerful nonlinear tools that can capture highly complex agricultural data relationships.

Applications in Agriculture:

- ✓ Predicting crop yields based on soil quality and weather conditions.
- ✓ Classifying disease outbreaks in crops using satellite images.
- ✓ Forecasting consumer demand for organic vs. conventional produce.

7. Nonlinear Error Correction Models (NECM)

NECMs extend traditional co-integration models by allowing for asymmetric adjustments when agricultural markets deviate from equilibrium.

Applications in Agriculture:

- ✓ Analyzing the long-term relationship between fertilizer prices and farm productivity.
- ✓ Studying the impact of policy interventions on agricultural supply chains.
- ✓ Modeling the nonlinear effects of exchange rates on agricultural trade.

8. Generalized Additive Models (GAMs)

GAMs allow for flexible relationships between variables by using smoothing functions rather than assuming strict linearity.

Applications in Agriculture:

- ✓ Modeling nonparametric relationships between temperature and crop yield.
- ✓ Estimating the effects of soil quality on land productivity.
- ✓ Analyzing consumer preferences for food products based on socioeconomic factors.

9. Bayesian Nonlinear Econometric Models

Bayesian approaches incorporate prior knowledge and update probabilities as new data becomes available, making them useful for decision-making under uncertainty.

Applications in Agriculture:

- ✓ Predicting future agricultural output under different climate change scenarios.
- ✓ Estimating the effectiveness of new farming technologies with limited data.
- ✓ Analyzing the probability of pest outbreaks based on historical trends.

10. Copula-Based Models

Copula models capture nonlinear dependencies between multiple variables, making them useful for studying interconnected agricultural risks.

- ✓ Modeling joint dependencies between weather patterns and crop yields.
- ✓ Understanding how price shocks in one agricultural market affect others.
- ✓ Assessing the combined impact of climate and economic risks on farm income.

Conclusion

Nonlinear econometric models are essential for capturing the complexities of agricultural systems. These models help researchers and policymakers better understand price volatility, climate change effects, and farm productivity dynamics, ultimately supporting more informed decision-making in the agricultural sector.

References

1. Gujarati, D. N., & Porter, D. C. (2009). *Basic Econometrics*. 5th edition. McGraw-Hill.
2. Wooldridge, J. M. (2016). *Introductory Econometrics: A Modern Approach*. 6th edition. Cengage Learning.
3. Greene, W.H. (2018). *Econometric Analysis*. 8th edition. Pearson.

Multivariate Analysis: Multivariate analysis refers to more than two variables being analyzed simultaneously where variables may be interrelated.

Multivariate Regression Analysis: Multivariate regression is a technique that estimates a single regression model with more than one interrelated outcome variable.

Multivariate Simple Regression Analysis: Multivariate regression is a technique that estimates a single regression model with more than one interrelated outcome variable where input variable is one.

Example 1: A survey was conducted to know the socio-economic condition of the labours in Bangabandhu Sheikh Mujib Safari Park from August, 2016 to November, 2016. Livelihood security index (LSI), Food security index (FDI), Health security index (HS_index), Household income (HH_in), BMI, Household size (HH_size), Sickness, Food convenient (food_convenient), Meals, Food group are considered as target variables. The data are given below-

LSI	FS_index	HS_index	HH_in	BMI	HH_size	sickness	food_convenient	meals	Food_group
0.358612	0.641758	0.31982	3501.389	22.78257	5	48	12	4	5
0.333837	0.327507	0.166268	4833.333	20.78291	5	10	12	3	4
0.311766	0.596034	0.207789	10833.33	21.479	2	36	12	3	8
0.335266	0.590841	0.293301	6683.333	18.34323	3	96	11	3	7
0.362136	0.357571	0.32903	3000	21.3499	8	72	12	3	5
0.286096	0.376572	0.337626	5123.333	21.3499	5	84	12	3	5
0.47378	0.738439	0.240182	4611.111	22.37874	6	24	12	5	8
0.458806	0.676662	0.371402	3400	25.68052	4	24	12	4	8
0.361683	0.774163	0.336419	7023.148	24.19656	9	60	10	5	9
0.432019	0.405507	0.262078	9152.778	23.08354	9	36	11	3	8
0.255191	0.513992	0.133509	8972.222	21.19145	3	6	11	3	7
0.393989	0.483496	0.374999	4166.667	26.78112	6	60	12	3	6
0.46822	0.643832	0.32926	6523.81	20.1125	7	60	12	5	6

0.380658	0.371587	0.208235	5138.889	21.52782	6	24	12	3	5
0.425844	0.644392	0.222905	4416.667	22.17746	6	24	12	4	8
0.334844	0.495836	0.168388	4190.476	21.94829	7	10	12	4	4
0.354085	0.517857	0.244233	6666.667	19.57075	6	48	12	4	6
0.380193	0.518771	0.26222	1318.182	27.62313	11	2	12	4	5
0.368037	0.574343	0.228991	5000	25.38428	2	7	12	4	5
0.361077	0.587378	0.17104	17500	23.47934	4	5	12	4	7
0.231452	0.460183	0.195625	2250	20.17756	4	4	12	3	5
0.255108	0.309782	0.276713	3875	25.38428	4	7	11	3	5
0.324019	0.611383	0.24724	6000	22.01188	3	5	12	3	6
0.152003	0.326603	0.155125	1800	20.71735	5	6	10	3	6
0.243872	0.509819	0.21125	2354.167	23.12906	4	5	12	3	5
0.25452	0.54228	0.152703	3500	20.02677	3	5	12	3	6
0.299512	0.428718	0.164028	1500	20.54442	5	4	12	3	6
0.252751	0.578782	0.12608	2875	19.07696	4	5	12	3	8
0.393805	0.643584	0.256277	5770.833	23.46196	4	5	12	4	6
0.373872	0.723211	0.168774	3888.889	19.52637	5	2	12	4	8
0.245064	0.485633	0.22364	2000	21.479	4	9	12	3	5
0.412703	0.548204	0.200228	4833.333	21.40793	5	5	12	3	8
0.363762	0.759001	0.18715	5000	18.745	3	5	12	4	9
0.459269	0.697394	0.134857	17052.5	17.79159	5	1	12	4	9
0.288056	0.674824	0.295088	7000	26.20973	2	8	12	3	8
0.251493	0.634397	0.104224	2250	20.17756	4	1	12	3	9
0.241044	0.62703	0.187598	3000	22.01188	3	5	12	3	7
0.277702	0.712062	0.209715	2300	20.82778	4	5	12	3	8
0.560036	0.665537	0.176879	7960.208	23.8462	8	1	12	4	9
0.358551	0.776579	0.249683	0	25.97572	3	5	12	4	8
0.508496	0.616068	0.213891	0	23.43164	11	2	12	4	8
0.339447	0.545388	0.212972	5416.667	24.9468	4	6	12	3	8
0.292207	0.564074	0.261681	8000	22.06157	2	10	12	3	6
0.320391	0.36115	0.340272	3950	24.135	6	60	12	3	4
0.344074	0.482024	0.334708	9166.667	24.9468	6	48	11	3	5
0.335873	0.525385	0.438457	10055.56	24.59722	4	84	12	3	3

0.358915	0.317681	0.575735	4700	17.79159	4	60	12	3	2
0.359541	0.352213	0.433455	2900	21.19145	5	48	12	3	3
0.472235	0.457531	0.362099	3250	24.90822	7	84	12	3	5
0.415888	0.49252	0.337273	1701.389	19.92658	6	24	12	3	7

Hands-on exercise of multivariate regression analysis using STATA software

Algorithm 1: (Multivariate simple regression model by Stata)

Step 1:

$Open\ Stata \rightarrow File \rightarrow \text{Im}\,port\ Data \rightarrow Excel\ Spreadsheet \rightarrow click\ or\ click\ Enter$

Step 2: In the Import Excel dialog box

- Browse Excel File
- Select Worksheet
- Tick Import first row as variable name

Click OK

Step 3:

$$Statistics \rightarrow Multivariateanalysis \rightarrow MANOVA, Multivariateandrelated \rightarrow MultivariateRegression$$

In “mvreg-Multivariate regression” dialog box-

- Select Dependent Variables
- Select Independent Variables

Click OK

The Stata output are given below-

```
. mvreg LSI FS_index HS_index = HH_size

Equation             Obs   Parms        RMSE    "R-sq"          F        P
--------------------------------------------------------------------------
LSI                   50       2     .0689675   0.2914   19.73694   0.0001
FS_index              50       2     .1302303   0.0163   .7958243   0.3768
HS_index              50       2     .0931668   0.0177   .8642821   0.3572

------------------------------------------------------------------------------
             |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
LSI          |
     HH_size |   .0207053   .0046606     4.44   0.000     .0113346    .0300761
       _cons |   .2444953   .0253478     9.65   0.000     .1935301    .2954606
-------------+----------------------------------------------------------------
FS_index     |
     HH_size |  -.0078509   .0088005    -0.89   0.377    -.0255456    .0098438
       _cons |    .584723    .047864    12.22   0.000      .488486      .68096
-------------+----------------------------------------------------------------
HS_index     |
     HH_size |   .0058531   .0062959     0.93   0.357    -.0068057    .0185119
       _cons |   .2234397   .0342419     6.53   0.000     .1545917    .2922877
------------------------------------------------------------------------------
```

❖ Multivariate Multiple Regression Analysis

Multivariate regression is a technique that estimates a single regression model with more than one interrelated outcome variable where input variables are more than one.

Algorithm 2: (Multivariate multiple regression model by Stata)

Step 1:

$$Open\ Stata \rightarrow File \rightarrow \mathrm{Im}port\ Data \rightarrow Excel\ Spreadsheet \rightarrow click\ or\ click\ Enter$$

Step 2: In Import Excel dialog box

- Browse Excel File
- Select Worksheet
- Tick Import first row as variable name

Click OK

Step 3:

$$\boldsymbol{Statistics \rightarrow Multivariateanalysis \rightarrow MANOVA, Multivariateandrelated \rightarrow MultivariateRegression}$$

In "mvreg-Multivariate regression" dialog box-

- Select Dependent Variables
- Select Independent Variables

Click OK

The Stata output are given below-

```
. mvreg LSI FS_index HS_index = HH_in BMI HH_size sickness food_convenient meals Food_group
```

Equation	Obs	Parms	RMSE	"R-sq"	F	P
LSI	50	8	.0487716	0.6899	13.35003	0.0000
FS_index	50	8	.0578927	0.8299	29.27454	0.0000
HS_index	50	8	.0579859	0.6670	12.02068	0.0000

	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
LSI						
HH_in	4.84e-06	2.12e-06	2.29	0.027	5.69e-07	9.11e-06
BMI	.0034771	.0029724	1.17	0.249	-.0025214	.0094756
HH_size	.0178419	.0038517	4.63	0.000	.0100689	.025615
sickness	.0007314	.0002915	2.51	0.016	.0001431	.0013197
food_convenient	.072035	.0151502	4.75	0.000	.0414606	.1026094
meals	.0379551	.0133744	2.84	0.007	.0109643	.0649458
Food_group	.0121842	.0047164	2.58	0.013	.0026661	.0217024
_cons	-.9203663	.20271	-4.54	0.000	-1.329452	-.511281
FS_index						
HH_in	-2.47e-06	2.51e-06	-0.98	0.331	-7.54e-06	2.60e-06
BMI	.0055977	.0035282	1.59	0.120	-.0015225	.012718
HH_size	-.0214581	.004572	-4.69	0.000	-.0306848	-.0122314
sickness	.0004708	.000346	1.36	0.181	-.0002275	.0011691
food_convenient	.0456889	.0179836	2.54	0.015	.0093966	.0819812
meals	.1157487	.0158757	7.29	0.000	.0837103	.1477871
Food_group	.0461411	.0055985	8.24	0.000	.0348429	.0574393
_cons	-.6999284	.2406199	-2.91	0.006	-1.185519	-.2143378
HS_index						
HH_in	-2.23e-06	2.52e-06	-0.89	0.381	-7.31e-06	2.85e-06
BMI	.0085734	.0035339	2.43	0.020	.0014417	.0157052
HH_size	-.0036831	.0045794	-0.80	0.426	-.0129247	.0055584
sickness	.002239	.0003466	6.46	0.000	.0015396	.0029384
food_convenient	.010987	.0180125	0.61	0.545	-.0253638	.0473377
meals	.0154608	.0159012	0.97	0.336	-.0166292	.0475508
Food_group	-.0137705	.0056075	-2.46	0.018	-.0250869	-.0024541
_cons	-.0592376	.2410074	-0.25	0.807	-.5456101	.427135

Reference:

Johnson, R. A. and Wichern, D. W. (2002). *Applied Multivariate Statistical Analysis*, 5th edition, Pearson Education.

Basic concepts of multivariate autoregressive model for forecasting agro-economic data

Prof. Dr. Sheikh Mohammad Sayem

A vector time series consists of multiple single time series observed simultaneously. While simple univariate ARMA models perform well for forecasting individual series, we often need multivariate models for the following reasons:

- To understand the dynamic relationships between multiple variables.
- To improve forecasting accuracy by incorporating information from related series.

The Vector AutoRegressive (VAR) model, popularized by Christopher Sims (1980) in *Macroeconomics and Reality*, extends univariate autoregression to a multivariate framework. It is widely used in empirical economics and financial modelling.

Practicing multivariate autoregressive model for forecasting agro-economic data using STATA software

Example: Stock Market Dynamics

Stock price fluctuations in one market (e.g., Dhaka Stock Exchange) can influence another market (e.g., Chittagong Stock Exchange). A joint dynamic model like VAR helps in understanding such interrelations and improving forecast accuracy.

Characteristics of VAR

- ✓ **Generalization of Univariate Models:** VAR extends the autoregression framework to multiple time series.
- ✓ **Endogeneity of Variables:** All variables are treated as endogenous, meaning there are no predetermined explanatory variables.
- ✓ **Lagged Dependence:** Each equation in the VAR model includes lagged values of all dependent variables, but no contemporaneous values (not existing or occurring at the same time as another).
- ✓ **Estimation:** Each equation can be estimated separately using Ordinary Least Squares (OLS), although Generalized Least Squares (GLS) provides better efficiency when error terms are correlated.

Mathematical Representation

Consider a bivariate system, where

$y_{1t}$is the inflation rate

$y_{2t}$is the unemployment rate

The relationship between them is captured by the Phillips Curve

A first-order VAR (VAR(1)) model is given by:

A VAR (1) in two variables can be written in matrix form (more compact notation) as

$$\begin{bmatrix} y_{1,t} \\ y_{1,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} + \begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} e_{1,t} \\ e_{2,t} \end{bmatrix}$$

in which only a single A matrix appears because this example has a maximum lag p equal to 1), or, equivalently, as the following system of two equations

$$y_{1,t} = c_1 + a_{1,1}y_{1,t-1} + a_{1,2}y_{2,t-1} + e_{1,t}$$

$$y_{2,t} = c_2 + a_{2,1}y_{1,t-1} + a_{2,2}y_{2,t-1} + e_{2,t}$$

Each variable in the model has one equation. The current (time t) observation of each variable depends on its own lagged values as well as on the lagged values of each other variable in the VAR.

Each variable depends on its own lagged value and the lagged value of the other variable.

Choosing the Lag Length in VAR

- ✓ The choice of lag length is crucial in estimating VAR models and performing Granger causality tests.
- ✓ Commonly, the lag length corresponds to the frequency of data:
  - → Quarterly data: 4 lags
  - → Monthly data: 12 lags
- ✓ Alternatively, optimal lag length can be determined using Information Criteria:
  - → Akaike Information Criterion (AIC)
  - → Schwarz-Bayesian Information Criterion (SBIC)
  - → Hannan-Quinn Criterion (HQIC)

Vector Autoregression (VAR)

1.Box–Jenkins and VAR approaches to economic forecasting are alternatives to traditional single- and simultaneous-equation models.

2.The VAR approach to forecasting considers several time series at atime. The distinguishing features of VAR are as follows:

- ✓ It is a truly simultaneous system in that all variables are regarded as endogenous.
- ✓ In VAR modelling the value of a variable is expressed as a linear functionof the past, or lagged, values of that variable and all other variables includedin the model.
- ✓ If each equation contains the same number of lagged variables in the system, it can be estimated by OLS without resorting to any systemsmethod, such as two-stage least squares (2SLS) or seemingly unrelated regressions(SURE).
- ✓ This simplicity of VAR modelling may be its drawback. In view of the limited number of observations that are generally available in most economic analyses, introduction of several lags of each variable can consume a lot of degrees of freedom.
- ✓ If there are several lags in each equation, it is not always easy to interpreteach coefficient, especially if the signs of the coefficients alternate.For this reason, one examines the impulse response function (IRF) in VARmodelling to find out how the dependent variable responds to a shock administeredto one or more equations in the system.
- ✓ There is considerable debate and controversy about the superiority ofthe various forecasting methods. Single-equation, simultaneous-equation,Box–Jenkins, and VAR methods of forecasting have their admirers as well asdetractors. All one can say is that there is no single method that will suit all situations. If that were the case, there would be no need to discussthe various alternatives. One thing is sure: The Box–Jenkins and VAR methodologies have now become an integral part of econometrics.

For example, three variables- GDP, investment and consumption from BBS. Whether these three variables are co-integrated we have to run Johanson co-integration test. It is to be noted that most of the time series data are non-stationary in the label form. But if we take the first difference or second difference they become stationary. First of all let us assume that they are non-stationary in the level form. To test co-integration we have to follow this algorithm.

Open STATA data from the STATA program or obtained from Excel.

Click statistics ⟶ Multivariate time series ⟶ Co-integrating rank of a VECM

```
. vecrank gdp consumption investment, trend(constant) max

                    Johansen tests for cointegration
Trend: constant                                 Number of obs =      10
Sample:  2008 - 2017                                     Lags =       2
------------------------------------------------------------------------
                                                     5%
maximum                                  trace    critical
  rank    parms       LL      eigenvalue  statistic    value
    0       12    -219.97126          .     41.0744    29.68
    1       17    -207.37922     0.91941    15.8903    15.41
    2       20     -201.2501     0.70648     3.6321*    3.76
    3       21    -199.43405     0.30456
------------------------------------------------------------------------
                                                     5%
maximum                                   max     critical
  rank    parms       LL      eigenvalue  statistic    value
    0       12    -219.97126          .     25.1841    20.97
    1       17    -207.37922     0.91941    12.2583    14.07
    2       20     -201.2501     0.70648     3.6321     3.76
    3       21    -199.43405     0.30456
------------------------------------------------------------------------
```

Here, the null hypothesis is that there is no Co-integrated equation among the three variables. Here, rank= o, 1, 2, 3 means no co-integrated equation, 1 co-integrated equation, 3 co-integrated equation. It appears from both the trace statistics and max statistics, they are greater than 5 percent critical value. So, the null hypothesis is rejected of no co-integration. That means they are integrated. And they move together in the long run. If they is no co-integrated equation we can run unrestricted VAR model. Since they are co-integrated we shall have to run Vector Error Correction Model (VECM). So, just to know the unrestricted VAR technique we assume that the three variables- GDP, consumption and income are not co-integrated.

Click statistics ⟶ Multivariate time series ⟶ Vector auto regression (VAR)

Vector autoregression

Sample: 2008 - 2017		Number of obs	= 10
Log likelihood =	-199.4341	AIC	= 44.08681
FPE =	7.68e+15	HQIC	= 43.38975
Det(Sigma_ml) =	4.22e+13	SBIC	= 44.72224

Equation	Parms	RMSE	R-sq	chi2	P>chi2
gdp	7	373.729	0.9977	4326.837	0.0000
consumption	7	2247.23	0.8493	56.37059	0.0000
investment	7	83.9678	0.9989	9082.565	0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
gdp						
gdp						
L1.	-1.007433	.3220233	-3.13	0.002	-1.638587	-.376279
L2.	-1.153967	.3244265	-3.56	0.000	-1.789831	-.5181025
consumption						
L1.	.1293952	.0337201	3.84	0.000	.0633051	.1954853
L2.	.0982085	.0334457	2.94	0.003	.0326562	.1637609
investment						
L1.	8.535973	2.07238	4.12	0.000	4.474184	12.59776
L2.	1.24857	1.42186	0.88	0.380	-1.538225	4.035366
_cons	2255.329	596.0245	3.78	0.000	1087.143	3423.516
consumption						
gdp						
L1.	-.5150686	1.936327	-0.27	0.790	-4.310199	3.280062
L2.	-3.522278	1.950777	-1.81	0.071	-7.345731	.3011756
consumption						
L1.	.2828373	.2027588	1.39	0.163	-.1145627	.6802372
L2.	.5560854	.2011091	2.77	0.006	.1619188	.950252
investment						
L1.	-38.59592	12.46122	-3.10	0.002	-63.01947	-14.17237
L2.	55.40518	8.54965	6.48	0.000	38.64817	72.16219
_cons	11411.18	3583.897	3.18	0.001	4386.875	18435.49
investment						
gdp						
L1.	.0010332	.0723508	0.01	0.989	-.1407717	.1428381
L2.	-.1059217	.0728907	-1.45	0.146	-.2487848	.0369415
consumption						
L1.	.0150162	.0075761	1.98	0.047	.0001674	.029865
L2.	.0075127	.0075144	1.00	0.317	-.0072153	.0222407
investment						
L1.	1.500825	.4656131	3.22	0.001	.5882397	2.41341
L2.	-.0811791	.3194574	-0.25	0.799	-.707304	.5449458
_cons	-74.02248	133.9122	-0.55	0.580	-336.4855	188.4405

Here, GDP, consumption and investment are dependent variables and their lag variables are independent variables. It appears from the table that most of the coefficients are significant because their p values are less than 5 percent. As per example Z value of GDP L1= -3.13 and p value is 0.002 which is less than 5 percent. So, the coefficient of GDP L1 is significant to explain GDP. Z value of investment L1is .01 and p value is 98.9 percent which is greater than 5 percent. That indicates that investment L1 has not any significant effect to explain investment.

In case of VAR model, all the coefficients are short run causalities. To run causality we shall have to run Granger causality test.

Click Statistics ⟶ Multivariate Time Series ⟶ VAR diagnostics and tests Granger Causality tests

Granger causality Wald tests

Equation	Excluded	chi2	df	Prob > chi2
gdp	consumption	23.705	2	0.000
gdp	investment	34.462	2	0.000
gdp	ALL	37.56	4	0.000
consumption	gdp	4.3206	2	0.115
consumption	investment	42.965	2	0.000
consumption	ALL	51.908	4	0.000
investment	gdp	3.3547	2	0.187
investment	consumption	4.9913	2	0.082
investment	ALL	8.3426	4	0.080

It appears from the table that most of the coefficients are significant since the p values are very small. As per example there is short run causality running from consumption to GDP. But In case of GDP there is no short term causality running from GDP to investment because the p value is 18.7 percent. If we consider both GDP and Consumption we see that p value is 0.000 that indicates there is short term causality running from GDP and consumption to investment.

Reference:

- Damodar N. Gujarati (2003), Basic Econometrics, 4th edition, McGraw Hill.
- Spyros Makridakis, Steven C. Wheelwright and Rob J. Hyndman (1998*),* *Forecasting Methods and Applications*, 3rd edition, John Wiley & Sons. Inc.

Research scope in socio-economic perspectives: Advancements and challenges in forecasting

Dr. Md. Shofiqul Islam
Principal Scientific Officer (PSO)
Agricultural Economics and Rural Sociology (AERS) Division
Bangladesh Agricultural Research Council, Farmgate, Dhaka
Email address: shafiqbau07@gmail.com
Mobile no.: 01704-778929

Agriculture remains the backbone of many economies, especially in developing nations where it provides employment and sustains livelihoods for millions. However, for agriculture to continue contributing effectively to poverty reduction and food security in the face of changing global dynamics, it must evolve. As technology advances, markets become more interconnected and policy frameworks shift, it is crucial to assess how these changes impact rural communities, food production, and economic development. This manual delves deeply into key areas of research in agriculture, focusing on market intelligence, policy research, innovative commercial agriculture, fallow land utilization and investment analysis of agricultural projects. By exploring these areas, this topic aims to guide stakeholders in designing inclusive, sustainable and economically viable agricultural strategies.

Socio-Economic Perceptions and Impact of Agricultural Technologies

The adoption of new agricultural technologies is a key driver of increased productivity and efficiency in the agricultural sector. However, the uptake of these technologies is heavily influenced by socio-economic perceptions and local contexts. This research focuses on understanding these perceptions and the impact of these technologies on rural communities.

Key Areas of Research:

- **Adoption Barriers**: Understanding the barriers that prevent farmers, particularly smallholders, from adopting new technologies. These include access to credit, training, perceived risk, and cultural resistance to change.
- **Technological Acceptance and Perceptions**: Investigating how farmers perceive the potential of agricultural technologies, including benefits such as increased yield, reduced labor, and profitability, and concerns like high costs, lack of support, and potential environmental impact.
- **Gender and Social Equity**: Exploring how different groups within rural communities especially women, youth, and marginalized groups experience the impacts of

technology adoption. Ensuring that technologies are inclusive and accessible for all socio-economic groups is critical.

Advancements:

- **Gender-Sensitive Technologies**: Ensuring that innovations are designed with gender considerations in mind, such as technologies that empower women through access to labor-saving devices or improved farming methods.
- **Participatory Research**: Involving farmers in the design and testing of technologies to ensure they meet the needs and preferences of end-users, ensuring better adoption rates.

Challenges:

- **Affordability**: Many new agricultural technologies are too expensive for small-scale farmers to afford, especially in developing countries.
- **Cultural Resistance**: Traditional farming practices may create resistance to adopting new technologies, especially if they are perceived as conflicting with local knowledge or cultural practices.

Efficient Production and Marketing Systems Development

Developing efficient agricultural production and marketing systems is essential to the long-term sustainability of agriculture. An efficient system reduces waste, ensures fair pricing, and improves farmers' access to profitable markets.

Key Areas of Research:

- **Market Linkages and Access**: Investigating mechanisms to link farmers directly with markets, cutting down intermediaries who often absorb a significant portion of the farmer's potential income. This includes exploring cooperatives, contract farming, and digital platforms.
- **Supply Chain Management**: Researching how to optimize the agricultural supply chain from production to market. This includes reducing transportation costs, improving storage facilities, and minimizing post-harvest losses.
- **Price Discovery and Market Transparency**: Ensuring farmers have access to accurate and timely market price information through market intelligence platforms which help them make informed decisions about when and where to sell their products.

Advancements:

- **Cold Chain Infrastructure**: The development of cold chain solutions especially in developing regions can help preserve perishable products reducing waste and extending market reach for farmers.
- **Digital Platforms and E-Commerce**: Digital solutions like online marketplaces, mobile apps, and e-commerce platforms that enable farmers to reach broader markets directly, eliminating middlemen, and enhancing profitability.

Challenges:

- **Infrastructure Gaps**: Many rural regions still lack the basic infrastructure necessary to support efficient marketing systems, including roads, storage facilities and reliable electricity.
- **Market Volatility**: Agricultural prices are often highly volatile due to seasonal changes, global demand shifts and climate change impacts which can affect farmers' income stability.

Market Intelligence

Market intelligence is the process of collecting, analyzing, and interpreting data about the market environment, consumer demand, competitive landscapes and pricing trends to make better business decisions. This research area helps farmers and agribusinesses align their strategies with current and future market trends.

Key Areas of Research:

- **Consumer Behavior Analysis**: Understanding consumer demand and preferences, including emerging trends such as organic, fair trade and sustainable products. This allows farmers to identify high-demand markets and align their production strategies accordingly.
- **Pricing Mechanisms and Forecasting**: Researching ways to improve price transparency through market intelligence tools and forecasting systems that help farmers plan better and reduce risk.
- **Global Market Trends**: Exploring international trade policies, shifts in global demand, and market entry strategies for farmers and agribusinesses, particularly in export-oriented agriculture.

Advancements:

- **Big Data and AI Integration**: The use of big data analytics and artificial intelligence (AI) to analyze trends and predict market movements, consumer preferences and pricing patterns, enabling farmers and agribusinesses to anticipate changes and make informed decisions.
- **Mobile Apps and SMS-Based Market Information**: Mobile applications that provide real-time updates on market prices, weather forecasts and other relevant data, empowering farmers with information at their fingertips.

Challenges:

- **Data Gaps**: Accessing accurate and timely market data can be difficult, especially in rural areas where internet and technology penetration is low.
- **Market Fragmentation**: In many regions, markets are fragmented and there is often a lack of coordination between smallholder farmers and other market participants, which makes data collection and sharing challenging.

Policy Research

Agricultural policies shape the environment in which farmers operate. Effective policy can promote sustainable practices, foster innovation and ensure equitable distribution of agricultural benefits. This research area focuses on understanding the constraints and opportunities in policy frameworks.

Key Areas of Research:

- **Constraints to Technology Adoption**: Identifying how government policies can either enable or restrict the adoption of innovative agricultural technologies particularly in terms of subsidies, land tenure and infrastructure support.
- **Input Access and Affordability**: Examining how policies can improve access to agricultural inputs such as seeds, fertilizers and machinery especially for smallholder farmers in rural areas.
- **Policy for Fair Pricing and Market Access**: Investigating policies that ensure fair pricing for agricultural products and improve farmers’ access to both domestic and international markets.

- **Research-Extension Linkages**: Studying the effectiveness of policies that link agricultural research institutions with extension services to ensure that farmers benefit from the latest technologies and innovations.

Advancements:

- **Agricultural Policy Frameworks**: Designing national and regional agricultural policies that support sustainable practices, market access and fair competition while addressing the specific needs of vulnerable populations.
- **Public-Private Partnerships (PPP)**: Fostering collaboration between governments, private enterprises and farmers to create policies that drive inclusive agricultural growth and innovation.

Challenges:

- **Policy Gaps and Inefficiency**: Many developing countries struggle with inefficient policy implementation, which leads to poor outcomes for farmers especially in terms of access to markets, fair pricing and input availability.
- **Global Trade Policies**: International trade agreements can either facilitate or hinder access to global markets depending on how they are structured.

Innovative Commercial Agriculture

Innovative commercial agriculture is about moving beyond subsistence farming to creating profitable agribusinesses that add value through processing, marketing and technological innovation. This area of research focuses on exploring new business models and fostering entrepreneurship in agriculture.

Key Areas of Research:

- **Value-Added Processing**: Exploring opportunities for farmers and agribusinesses to process raw agricultural products into higher-value items. This could include processing fruits into juices or grains into flour which increases profit margins and job creation.
- **Agri-Entrepreneurship and Business Models**: Identifying sustainable and innovative business models that can be scaled and replicated across different regions. This includes contract farming, cooperatives and franchising.

- **Agribusiness Incubators and Support**: Developing agricultural incubators and accelerators that support entrepreneurs with the resources, mentorship and access to financing needed to succeed in commercial agriculture.

Advancements:

- **Agri-Tech Innovation**: Encouraging the development of agri-tech solutions that enhance production efficiency, reduce waste and increase sustainability such as precision farming and smart irrigation systems.
- **Contract Farming and Rural Industries**: Scaling up contract farming and establishing rural industrial clusters that process local agricultural products providing both market access and added value.

Challenges:

- **Capital Constraints**: Securing funding for innovative agribusinesses is often difficult especially for small-scale entrepreneurs due to high risk and long repayment periods.
- **Market Competition**: Smaller agribusinesses often face stiff competition from larger companies which dominate the value chains and markets.

Fallow Land Utilization Research

Fallow land, often left idle after years of farming, can be a valuable resource if managed correctly. This research focuses on finding productive uses for fallow lands to improve soil health and increase productivity.

Key Areas of Research:

- **Soil Restoration Techniques**: Researching techniques to restore the fertility of fallow lands, including agroforestry, crop rotation and organic farming practices.
- **Alternative Uses for Fallow Land**: Identifying non-cultivation alternatives for fallow land, such as grazing or growing bioenergy crops which could help farmers generate income while maintaining land health.
- **Sustainable Land Management**: Studying the long-term ecological and economic impacts of different land management practices to optimize fallow land use.

Advancements:

- **Agroecological Practices**: Encouraging the use of sustainable agricultural practices such as agroecology and agroforestry that integrate both food production and environmental conservation.
- **Precision Agriculture**: Using technology to monitor soil conditions and optimize the use of fallow land, ensuring it is restored to its full productivity potential.

Challenges:

- **Economic Viability**: Convincing farmers to invest in fallow land restoration can be challenging due to the upfront costs and the time required to see results.
- **Land Tenure Issues**: Many farmers are hesitant to invest in land restoration if they do not have secure land tenure or face uncertain land rights.

Investment Analysis of Agricultural Projects

Investment analysis is critical for understanding the financial viability of agricultural projects. Proper investment evaluation ensures that resources are allocated efficiently and that projects are sustainable in the long term.

Key Areas of Research:

- **Cost-Benefit Analysis**: Conducting detailed cost-benefit analysis of agricultural projects, taking into account factors like capital investment, operational costs, market conditions and risk.
- **Risk Management and Financial Products**: Developing risk management strategies for agricultural projects including insurance and hedging instruments to mitigate weather-related and market-related risks.
- **Impact Investment**: Studying how to attract impact investors to the agricultural sector with an emphasis on both financial returns and social/environmental benefits.

Advancements:

- **Innovative Financing Models**: The development of financial products such as weather-indexed insurance or blended finance mechanisms that provide capital for agricultural projects while addressing risks.

- **Agri-Financing Platforms**: Building platforms that connect agricultural projects with potential investors and lenders helping farmers access necessary funding.

Challenges:

- **Capital Access**: Farmers, particularly smallholders face challenges in accessing financing due to the perceived riskiness of agriculture.
- **Long Investment Horizon**: Agriculture is a long-term investment requiring patience and careful planning to manage risks and rewards effectively.

To merge the elements of market intelligence, policy research, innovative commercial agriculture, fallow land utilization and investment analysis, forecasting plays a pivotal role in guiding decisions at all levels of the agricultural sector. A well-structured forecasting system enables farmers, policymakers, agribusinesses, and investors to make informed decisions that maximize productivity, optimize resource allocation, and reduce risks associated with market volatility and environmental changes.

Key Takeaways:

- Accurate market forecasting helps in demand planning, pricing strategies and mitigating risks associated with market volatility.
- Policy forecasting ensures that agricultural systems are prepared for future changes in regulations, climate adaptation strategies and international trade policies.
- Technological advancements particularly in agri-tech allow for more precise forecasting of production outcomes, resource use and environmental impacts.
- Fallow land utilization forecasting helps optimize land use practices and predict long-term environmental and economic benefits.
- Investment analysis forecasting aids in predicting the financial viability of agricultural projects and improving access to finance reducing the risk for investors and farmers.

By improving market intelligence, formulating effective policies, fostering innovative commercial agriculture, utilizing fallow land and conducting comprehensive investment analysis, we can support sustainable agricultural growth that benefits all stakeholders. Moving forward, these research areas should be prioritized and refined to create resilient, profitable and equitable agricultural systems.