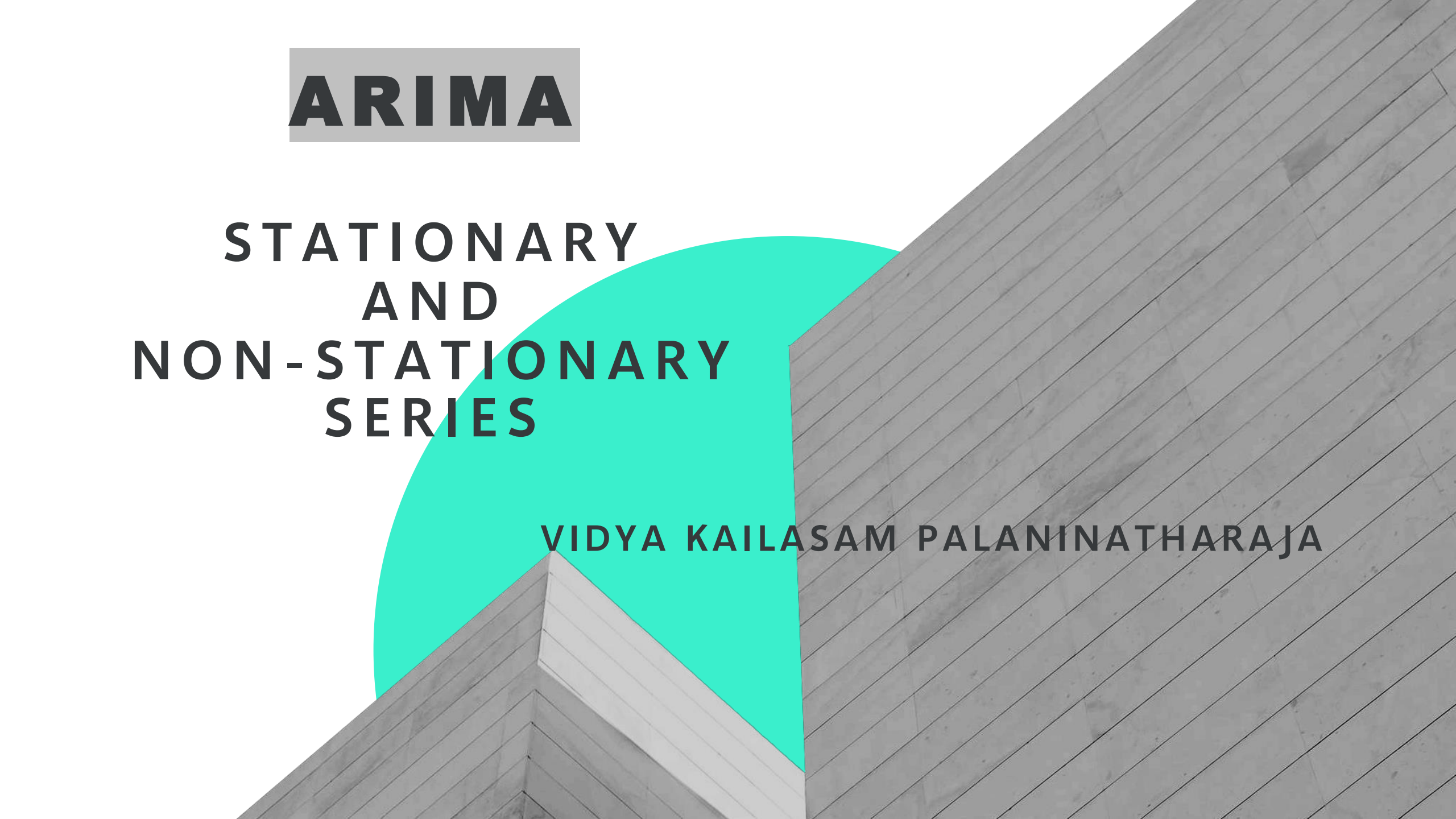
ARIMA

STATIONARY AND NON-STATIONARY SERIES

VIDYA KAILASAM PALANINATHARAJA



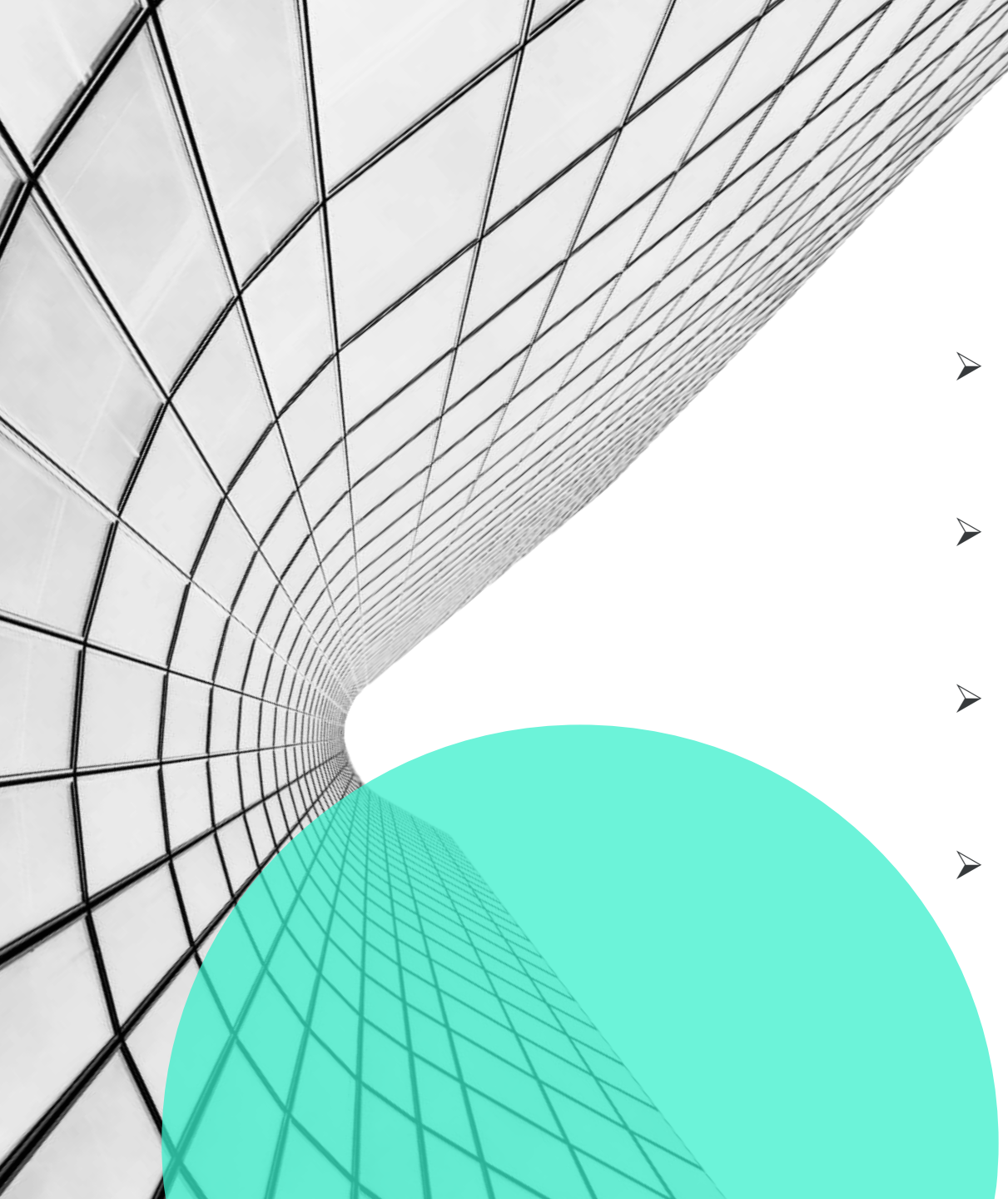
INTRODUCTION

- ARIMA (Autoregressive Integrated Moving Average) is a popular time series forecasting method used in Business Forecasting
- Components of ARIMA models:
  - Autoregression (AR) – past values (p)
  - Integration(I) – differences (d)
  - Moving Average(MA) – past errors (q)

BUILDING BLOCKS OF ARIMA MODEL

- Analyze the time series data
- Identify any trends or seasonality
- Determine the appropriate values for the ARIMA parameters (p, d, and q)
- Fit the model to the data and generate forecasts

✓ The accuracy of the forecasts can be evaluated using metrics such as Mean Absolute Error (MAE) or Root Mean Squared Error (RMSE)



CONCEPT OF STATIONARITY

- Stationary time series- statistical properties remain constant over time
- Nonstationary time series-statistical properties change over time
- Volatile business economy- nonstationary models are preferred->differencing
- Forecasts made one or two periods ahead may differ little between stationary and non-stationary models, but long-term trends and associated prediction intervals become very different

NETFLIX SALES SERIES

Table 6.5 MSE and RMSE for Different ARIMA Models for the Log Transformed Netflix Sales Series

(A) Results using Minitab

Model	MSE × 100	RMSE × 10	DF
AR(1), AR(2)	Cannot be fitted; violates stationarity assumption		
MA(1)	62.00	7.87	62
MA(2)	17.13	4.14	61
ARIMA(0,1,1)+C	0.447	0.67	61
ARIMA(1,1,0)+C	0.365	0.60	61
ARIMA(1,1,1)+C	0.370	0.61	60

(B) Results using R

Model	MSE × 100	RMSE × 10
MA(1)	63.972	7.998
MA(2)	18.939	4.352
ARIMA(0,1,1)+C	0.454	0.674
ARIMA(1,1,0)+C	0.386	0.621
ARIMA(1,1,1)+C	0.388	0.623

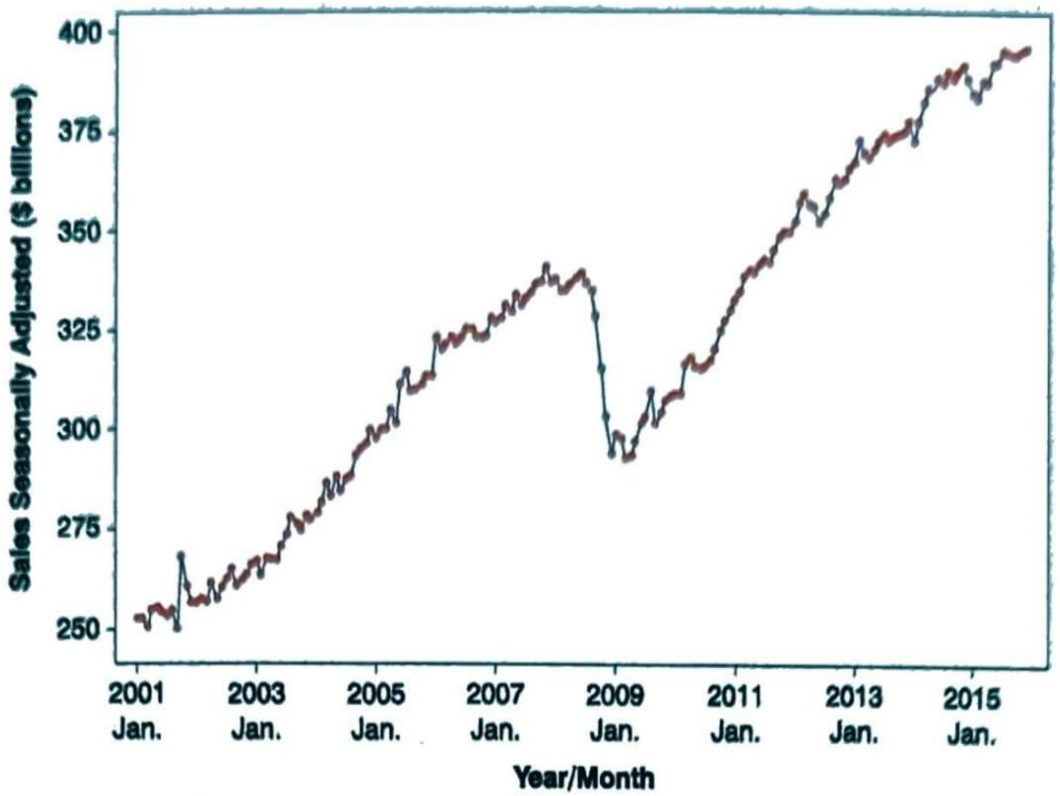
Data: Netflix_2.xlsx

NONSTATIONARY SERIES:
FITTED MODEL FOR NETFLIX DATA

- **MSE and RMSE – Evaluation metrics**
- **AR (1), AR (2) – Stationary – no fit**
- **Best-fit model ARIMA (1, 1, 0) + C**
- **ARIMA models having state space (exponential smoothing equivalents)**

US RETAIL SERIES

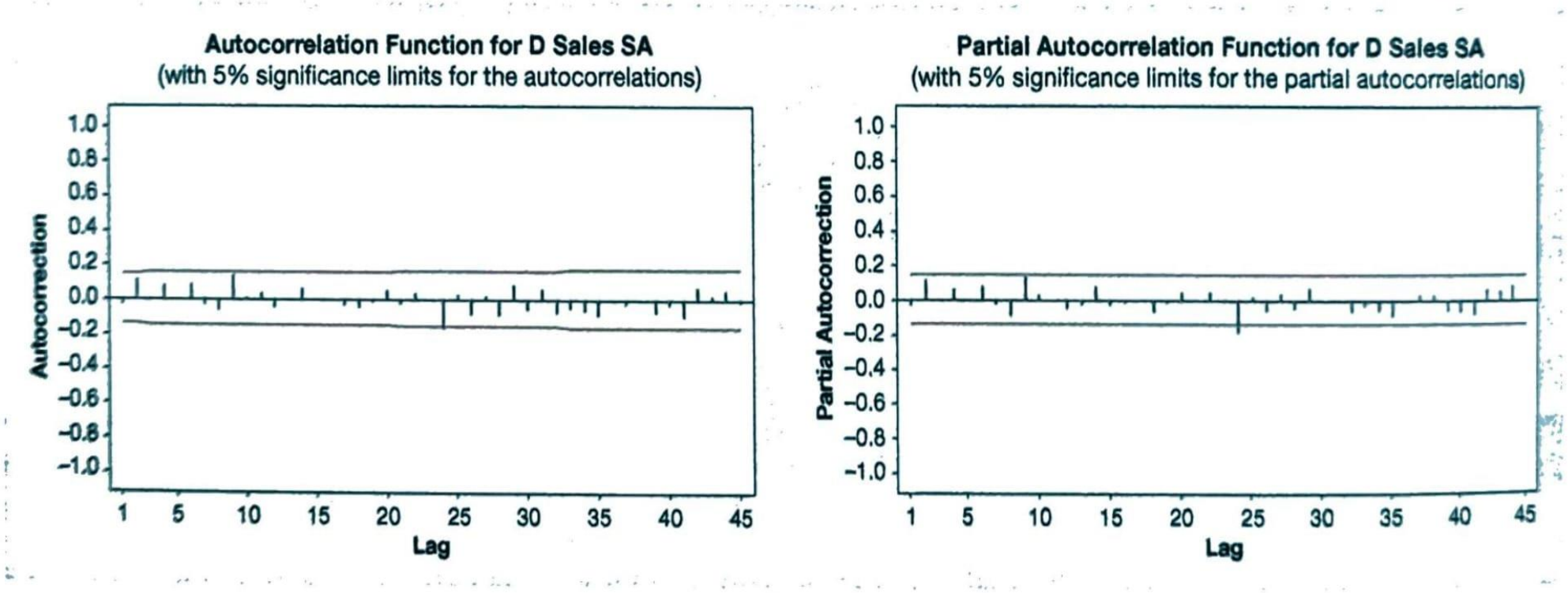
Figure 1.3 Seasonally Adjusted Series for U.S. Monthly Retailers Sales



Data: US_retail_sales_2_SA.xlsx

FIRST DIFFERENCE: ACF AND PACF

Figure 6.16 *ACF* (and *PACF*) for First Differences of U.S. Retail Sales *(seasonally adjusted)*



Data: US_retail_sales_2_SA.xlsx

NONSTATIONARY SERIES:
FITTED MODEL FOR U.S. RETAIL SALES

- Long time-series; Strong upward trend
- ACF and PACF
- ACF and PACF for first differences of U.S. retail sales
- Suggested models: AR(1), MA(1)
- Impact of great recession running from December 2007 to June 2009
- Inclusion of constant term

Table 6.6 Summary Results for ARIMA Models for U.S. Retail Sales

ARIMA Model	MSE [to 06/08]	MSE [to 12/15]
(1,1,0)	12.02	13.27
(1,1,0)+C	10.34	12.69
(0,1,1)	12.22	13.27
(0,1,1)+C	9.67	12.69
(2,1,0)+C	10.16	12.60
(0,1,2)+C	9.78	12.61
(1,1,1)+C	9.78	12.65
(2,2,0)	13.47	16.30

Data: US_retail_sales_2_SA.xlsx

RESULTS OF RECESSION

Pre-recession data

- ARIMA (0,1,1) + C preferred
- ARIMA (2, 2, 0) model is inferior

Post-recession data

- ARIMA (0, 1, 1) model- adequate
- MA term not significant

Final Estimates of Parameters

Period : 01/01–06/08

Type	Coefficient	Standard Errors	T-test	P-values
MA(1)	0.5648	0.0886	6.38	0.000
Constant	0.9803	0.1441	6.80	0.000

Period : 01/01–12/15

Type	Coefficient	Standard Errors	T-test	P-values
MA(1)	0.0307	0.0751	0.41	0.684
Constant	0.8003	0.2581	3.10	0.002

SUMMARY

- Random walk + drift = adequate model
- Random component->unpredictable fluctuations
- Drift component->long term trend

  REASON: The recession had a huge impact on the series and such effects typically lead to models that adapt rapidly to the latest information

- Differencing and decomposition