SUBJECT: FOUNDATIONS FOR BUSINESS ANALYTICS

Module 4: Time Series & Index Numbers

Exhaustive Applied Edition • Detailed Explanations with Solved Problems (Units 18 – 22)

18 Meaning and Significance – Utility, Components of Time Series - Measurement of Trend: Method of Least Squares

Meaning and Significance of a Time Series

A **Time Series** is a chronological sequence of quantitative data observations recorded at uniform, successive intervals over a specified period (e.g., daily stock closing prices, monthly retail revenues, or annual population metrics). In business analytics, a time series models dynamic environmental changes, allowing data scientists to isolate historical patterns, evaluate corporate performance velocity, and forecast future operations under uncertainty.

The Utility of Time Series Analysis

- **Future Forecasting:** Builds predictive models to forecast upcoming metrics like inventory needs, cash flows, and resource demand.
- **Performance Evaluation:** Compares actual performance indicators against historical baselines to detect system structural drifts.
- **Economic Cycle Auditing:** Isolates seasonal fluctuations and macro cyclical shifts to assist strategic planning.

The Four Core Components of a Time Series

A raw chronological data line is a composite mixture of four distinct underlying operational forces:

1. **Secular Trend (T):** The long-term smooth progression or direction of the data over an extended horizon (e.g., decades), reflecting baseline changes like population growth or technological adoption. Trends can be linear (straight line) or non-linear (curved).
2. **Seasonal Variation (S):** Short-term, highly rhythmic fluctuations that repeat within a fixed, predictable period of one year or less. Driven primarily by weather cycles, holidays, or institutional calendars (e.g., air conditioner sales spiking in summer, or retail spikes during Diwali/Christmas).
3. **Cyclical Variation (C):** Long-term wave-like oscillations that repeat over periods longer than a year, driven by systemic macroeconomic business cycles (recession, recovery, prosperity, depression). Unlike seasonal variations, cycles vary in length and intensity.
4. **Irregular / Random Variation (I):** Unpredictable, erratic residual fluctuations caused by random, unique shocks outside normal control configurations (e.g., natural acts, labor strikes, geopolitical crises, or pandemics).

Mathematical Structural Models

To analyze individual components, data scientists assemble them using two primary structural assumptions:

- **Additive Model:** Assumes components operate independently.
  $Y = T + S + C + I$
- **Multiplicative Model:** Assumes components interact proportionally, commonly used in business datasets.
  $Y = T \times S \times C \times I$

Measurement of Trend: The Method of Least Squares (Linear Trend)

The Method of Least Squares fits a straight trend line through a chronological data series in a way that minimizes the sum of the squared vertical differences (residuals) between actual data points and the fitted line.

> ***LINEAR TREND EQUATION***
>
> $Y = a + bX$

Component	*Description*
Y	*The calculated trend value for a target period.*
X	*The coded time variable coordinate centered around an origin point.*
a	*The intercept value representing the trend height at the designated origin point ($X = 0$).*
b	*The growth rate slope, indicating the average unit change in* ***Y*** *per unit shift in time.*

The Simplified Estimation Matrix (Coded Time Method)

By shifting the time origin to the exact mid-point of the series, the sum of the time codes is set to zero ($\Sigma X = 0$). This simplifies the normal equations into direct parameter estimation formulas:

$$a = \Sigma Y / n \quad | \quad b = \Sigma XY / \Sigma X^2$$

EXAMPLE PROBLEM & SOLUTION

Problem: An e-commerce platform tracks its annual revenue metrics over $n = 5$ consecutive years. Find the linear trend line equation using the Method of Least Squares, and forecast the projected revenue for the year 2026.

Year	Revenue (Y in ₹ Crores)
2021	10
2022	12
2023	15
2024	18
2025	20

Solution Strategy: Set the origin to the middle year (2023) so that $\Sigma X = 0$. Construct the computational matrix:

Year	Y	Coded X (Year – 2023)	X^2	XY
2021	10	–2	4	–20
2022	12	–1	1	–12
2023	15	0	0	0
2024	18	1	1	18
2025	20	2	4	40
Total (n=5)	**ΣY = 75**	**ΣX = 0**	**$\Sigma X^2 = 10$**	**ΣXY = 26**

1. *Calculate Parameter a: $a = 75 / 5 = 15$*
2. *Calculate Parameter b: $b = 26 / 10 = 2.6$*
3. *Construct Trend Line Equation:* $\mathbf{Y = 15 + 2.6X}$ *(Origin: 2023)*

Forecasting for 2026: Calculate the coded time coordinate for 2026: $X = 2026 - 2023 = 3$. Substitute $X = 3$ into the trend model:

$$Y_{2026} = 15 + 2.6(3) = 15 + 7.8 = \text{₹}\mathbf{22.8\ Crores}$$

19 Parabolic Trend and Logarithmic Trend

When long-run corporate growth metrics display significant curvature rather than constant linear changes, analytics managers deploy non-linear trend structures:

I. The Parabolic Trend (Second-Degree Polynomial Model)

Used to model data that experiences a changing rate of growth or decline over time, forming a U-shaped or inverted U-shaped curve.

PARABOLIC TREND FORMULA

$$Y = a + bX + cX^2$$

Where c measures the rate of acceleration or deceleration in the trend line curve.

When the time variable is coded so that $\Sigma X = 0$ and $\Sigma X^3 = 0$, the parameters are isolated using simplified system equations:

$$b = \Sigma XY / \Sigma X^2 \quad | \quad \text{Solve simultaneously: } \Sigma Y = na + c\Sigma X^2 \text{ and } \Sigma X^2 Y = a\Sigma X^2 + c\Sigma X^4$$

EXAMPLE PROBLEM & SOLUTION

Problem: A manufacturing plant records its production volumes, displaying non-linear growth traits. Given $n = 5$, coded time coordinates yield the following values: $\Sigma Y = 50, \Sigma X^2 = 10, \Sigma XY = 20, \Sigma X^2 Y = 110, \Sigma X^4 = 34$. Find the parameters and construct the parabolic trend model.

Solution Strategy: Substitute values into the simplified parameter equations to solve for $a, b,$ and c:

1. Calculate Parameter b: $b = 20 / 10 = \mathbf{2}$

2. Set Up the Simultaneous Equations for a and c:
Equation (i): $50 = 5a + 10c \rightarrow$ *Multiply by 2:* $100 = 10a + 20c$
Equation (ii): $110 = 10a + 34c$

3. Subtract Equation (i) from (ii) to Isolate c:
$(110 - 100) = (34c - 20c) \rightarrow 10 = 14c \rightarrow c = 10 / 14 = \mathbf{0.714}$

4. Substitute c back into Equation (i) to Isolate a:
$50 = 5a + 10(0.714) \rightarrow 50 = 5a + 7.14 \rightarrow 5a = 42.86 \rightarrow a = \mathbf{8.572}$

5. *Construct Parabolic Equation:* $Y = 8.572 + 2X + 0.714X^2$

II. The Logarithmic / Exponential Trend Model

Used when data experiences constant percentage growth over time rather than a constant absolute increase, resulting in an exponential curve (e.g., tech user compounding growth curves).

EXPONENTIAL MODEL FRAMEWORK

$$Y = a \cdot b^X \rightarrow \text{converted via logarithms to linear form: } \log(Y) = \log(a) + X \cdot \log(b)$$

By letting $Y^* = \log(Y)$, $A = \log(a)$, and $B = \log(b)$, the model becomes a linear equation: $Y^* = A + BX$. Under the coded time condition where $\Sigma X = 0$, the parameters are calculated as:

$$A = \Sigma Y^* / n \quad | \quad B = \Sigma XY^* / \Sigma X^2 \quad | \quad \text{Extracting: } a = \text{antilog}(A) \text{ and } b = \text{antilog}(B)$$

EXAMPLE PROBLEM & SOLUTION

Problem: A technology startup models its subscriber growth. After transforming user counts to base-10 logarithms ($Y^* = \log Y$) for $n = 5$ periods with $\Sigma X = 0$ and $\Sigma X^2 = 10$, the team calculates: $\Sigma Y^* = 10$ and $\Sigma XY^* = 3$. Find the exponential growth model values.

Solution Strategy: Calculate the log parameters A and B, then apply antilogs to find the original scale constants:

1. *Calculate A:* $A = 10 / 5 = 2.0000 \rightarrow a = \text{antilog}(2) = 10^2 = \mathbf{100}$ *(Base baseline users)*
2. *Calculate B:* $B = 3 / 10 = 0.3000 \rightarrow b = \text{antilog}(0.3) = 10^{0.3} = \mathbf{2}$ *(Growth multiplier factor)*
3. *Construct Original Exponential Growth Function:* $Y = 100 \cdot (2)^X$

Interpretation: The platform starts with a base trend of 100 subscribers at the origin point, and the user base doubles ($b = 2$) with every unit step forward in time.

20 Index Numbers: Meaning and Significance

Meaning of Index Numbers

An **Index Number** is a specialized statistical measure designed to calculate the relative percentage change in the price, quantity, or value of a variable or a basket of related items over time, relative to a chosen baseline period known as the **Base Year** (denoted with subscript 0). Index numbers function as economic indicators, simplifying complex multi-item price movements into a single tracking value.

Significance and Management Utility

- **Purchasing Power Measurement:** Used to track inflation, measure cost-of-living adjustments, and update salary matrices to preserve real consumer purchasing power.
- **Economic Trend Analysis:** Serves as a key indicator for tracking high-level economic changes, such as the Consumer Price Index (CPI) and Wholesale Price Index (WPI).
- **Deflating Nominal Series:** Converts nominal economic metrics into real, inflation-adjusted metrics to allow for accurate performance comparisons over time.

21 Problems in Construction of Index Numbers, Methods of Constructing Index Numbers & Tests of Adequacy

Systemic Problems Incurred in Index Construction

Designing a reliable index framework requires addressing several structural and statistical challenges:

- **Selection of the Base Year:** The base year must represent a stable economic period free from extreme shocks, natural acts, or high inflation anomalies. It should also not sit too far in the past, as consumer preferences shift over time.
- **Selection of the Commodity Basket:** Selecting a representative basket of goods that accurately captures the consumption habits of the target group.
- **Choice of Inherent Weights:** Determining the relative importance of each item in the basket. For example, a change in the price of food staples impacts the cost of living significantly more than a change in luxury goods pricing.

- **Choice of Average:** Deciding whether to use the arithmetic mean or the geometric mean (which is mathematically superior for tracking ratios but more complex to compute).

Methods of Constructing Price Index Numbers

Index numbers are classified based on how weighting constraints are applied to the commodities basket:

I. Unweighted Aggregative Price Index Method

A simple index structure that sums prices across items without adjusting for their relative economic importance or consumption volumes:

$$P_{01} = [\ \Sigma P_1 / \Sigma P_0\ ] \times 100$$

Where P_0 is the base year price, and P_1 is the current year price.

II. Weighted Aggregative Price Index Frameworks

Weighted index methods incorporate commodity consumption volumes (quantities) into the formula to ensure the index accurately reflects economic reality. We analyze three major classical formulas:

1. LASPEYRES PRICE INDEX (BASE-YEAR WEIGHTED)

$$P_L = [\ \Sigma P_1 Q_0 / \Sigma P_0 Q_0\ ] \times 100$$

Uses base year quantities (Q_0) as weights. It can be calculated cost-effectively because it does not require collecting new quantity data every year, but it tends to overstate inflation because it ignores consumer shifts toward cheaper alternatives.

2. PAASCHE PRICE INDEX (CURRENT-YEAR WEIGHTED)

$$P_P = [\ \Sigma P_1 Q_1 / \Sigma P_0 Q_1\ ] \times 100$$

Uses current year quantities (Q_1) as weights. It incorporates real-time consumer habits but requires continuous, expensive consumption surveys, and it tends to understate inflation.

3. FISHER'S IDEAL INDEX MODEL

$$P_F = \sqrt{[P_L \times P_P]} = \sqrt{[[\Sigma P_1Q_0 / \Sigma P_0Q_0] \times [\Sigma P_1Q_1 / \Sigma P_0Q_1]]} \times 100$$

Calculated as the geometric mean of the Laspeyres and Paasche indexes. It balances the biases of both formulas and is called "ideal" because it satisfies the primary mathematical adequacy tests.

EXAMPLE PROBLEM & SOLUTION MATRIX

Problem: Given the following budget data for two commodities (A and B), calculate the Laspeyres, Paasche, and Fisher Price Index Numbers for the current year.

Item	Base Year (0)		Current Year (1)	
	Price (P_0)	Quantity (Q_0)	Price (P_1)	Quantity (Q_1)
A	10	5	12	6
B	20	2	25	2

Solution Strategy: Build a cross-multiplication worksheet table to calculate the four baseline summation components required for the formulas:

Item	P_0Q_0	P_1Q_0	P_0Q_1	P_1Q_1
A	10 × 5 = 50	12 × 5 = 60	10 × 6 = 60	12 × 6 = 72
B	20 × 2 = 40	25 × 2 = 50	20 × 2 = 40	25 × 2 = 50
Sum (Σ)	**$\Sigma P_0Q_0 = 90$**	**$\Sigma P_1Q_0 = 110$**	**$\Sigma P_0Q_1 = 100$**	**$\Sigma P_1Q_1 = 122$**

Formula Multiplications:

1. Laspeyres Index Calculation:

$$P_L = (110 / 90) \times 100 = 1.2222 \times 100 = 122.22$$

2. Paasche Index Calculation:

$$P_P = (122 / 100) \times 100 = 1.2200 \times 100 = 122.00$$

3. Fisher's Ideal Index Calculation:

$P_F = \sqrt{[\ 122.222 \times 122.00\ ]} = \sqrt{14911.08} = 122.11$

Interpretation: The composite market price for this basket of commodities has increased by **22.11%** relative to the base year level, calculated using the balanced Fisher model.

Mathematical Tests of Adequacy for Index Numbers

To verify the mathematical consistency of an index formula, it is evaluated against standard structural tests:

1. **Time Reversal Test (TRT):** Asserts that if the base year and current year parameters are interchanged, the newly generated index should be the exact reciprocal of the original index. The test condition requires:

   $$P_{01} \times P_{10} = 1 \text{ (Omitting the 100 multiplier)}$$

   Fisher's Ideal Index satisfies this test, whereas the Laspeyres and Paasche formulas do not.

2. **Factor Reversal Test (FRT):** Asserts that interchanging the price variables (P) and quantity variables (Q) within the index formula should yield the true total value ratio of the basket. The test condition requires:

   $$P_{01} \times Q_{01} = \Sigma P_1Q_1 / \Sigma P_0Q_0$$

   Fisher's Ideal Index is one of the unique models that successfully satisfies this test, validating its theoretical consistency.

22 Chain Index Numbers

Fixed-Base vs. Chain-Base Index Models

In a **Fixed-Base Index**, the comparison baseline period remains static and locked to a single historical year over time. While useful for long-term tracking, it becomes less accurate if new products enter the market or old items vanish. Conversely, a **Chain-Base Index** uses the *immediately preceding year* as its base for each successive period. This creates a link of year-over-year comparisons, allowing the index to adapt to shifting consumer habits dynamically.

The Chain Index Conversion Algorithm

To construct a chain index, you first compute the year-over-year percentage change for each period, known as the **Link Relative (LR)**:

> ***Link Relative (LR) for Current Year = [Price of Current Year / Price of Previous Year] × 100***

These link relatives are then systematically linked together to form the continuous **Chain Index (CI)** series using the conversion formula:

> ***CHAIN INDEX CONVERSION EQUATION***
>
> ***Current Year's CI = [(Current Year's Link Relative × Previous Year's CI) / 100]***

EXAMPLE PROBLEM & SOLUTION

Problem: Given the pre-calculated annual Link Relatives for a commodity across three sequential years, construct the matching Chain Index numbers, setting the initial year baseline to 100.

- Year 2023: Link Relative = 100 (Initial base year)
- Year 2024: Link Relative = 110
- Year 2025: Link Relative = 120

Solution Strategy: Apply the Chain Index conversion formula sequentially from year to year:

Year 2023 (Base Line):

The initial index is set to **100**.

Year 2024 Chain Index Calculation:

CI_{2024} = [(Current LR × Previous CI) / 100] = [(110 × 100) / 100] = 110

Year 2025 Chain Index Calculation:

CI_{2025} = [(Current LR × Previous CI) / 100] = [(120 × 110) / 100] = [13200 / 100] = 132

Final Index Output Table:

Year	Calculated Link Relative (LR)	Compiled Chain Index (CI)
2023	100	**100.00**
2024	110	**110.00**
2025	120	**132.00**

Interpretation: By 2025, the cumulative price has increased by **32.00%** relative to the initial 2023 baseline, calculated through the year-over-year chained sequence framework.

End of Module 4 • Subject: Foundations for Business Analytics