Freebooks.pk

Free Notes • Past Papers • Guides for Students

ICS Part-II Statistics — Punjab Board

CHAPTER 16

Analysis of Time Series

Statistics • 2nd Year • ICS • Punjab Board • Free PDF Notes

Chapter 16: Analysis of Time Series

Almost every business, economic, or scientific measurement recorded repeatedly over time -- annual wheat production, monthly sales, hourly temperature -- forms a time series. Unlike the cross-sectional data considered in earlier chapters, a time series has an inherent ORDER: the position of an observation in time is itself meaningful information, not just another label. This chapter develops the standard framework for describing and decomposing such data: the four classical components (secular trend, seasonal variations, cyclical fluctuations, and irregular movements), the two standard models (multiplicative and additive) that combine them, and four practical methods for isolating and measuring the most important of these components, the secular trend -- free hand curve, semi-averages, moving averages, and least squares.

The mathematics of secular trend estimation by least squares directly reuses the machinery built in Chapter 14 for fitting a straight line, applied here with time itself as the independent variable, plus an extension to fitting a quadratic (parabolic) curve when a straight line does not adequately describe the pattern. A recurring practical theme is the CODING of the time variable -- shifting and rescaling the origin so that sum(x)=0, which dramatically simplifies the arithmetic of every method in the chapter.

Learning Objectives

- Define a time series and distinguish it from cross-sectional data
- Construct a historigram and identify the signal and noise components of a series
- Identify and describe the four classical components of a time series: secular trend, seasonal variations, cyclical fluctuations, and irregular movements
- Distinguish the multiplicative model from the additive model for combining time series components
- Code the time variable to simplify trend calculations
- Estimate secular trend using the method of free hand curve
- Estimate secular trend using the method of semi-averages

- Compute k-period (and centred) moving averages, and use them to smooth a time series
- Fit a linear and a quadratic secular trend using the method of least squares
- Shift the origin of an already-fitted trend equation to a new reference time period

Key Concepts

16.1 Time Series

A time series is the sequence y1, y2, ..., yn of n observations of a variable Y, recorded according to their time of occurrence t1, t2, ..., tn. Symbolically Y is expressed as a function of time, $y=f(t)+e$, where $f(t)$ is the systematic, determined component of variation (the signal) and e is the random, irregular component (the noise). Observations in a time series are usually recorded at equidistant points in time -- hourly temperature, annual wheat yield, monthly sales, daily attendance.

A historigram is a graph of a time series: time t is plotted on the x-axis, the observed values of Y on the y-axis, and successive points are joined by line segments. Constructing a historigram is always the first step in examining a time series, before any formal analysis is attempted.

16.2 Components of a Time Series

A time series may be composed of up to four distinct components, each the result of a well-defined but different underlying cause; a given series need not contain all four. The secular (long-term) trend, T, is a smooth, steady, gradual movement of the series in the same direction (up or down) over a fairly long period, reflecting forces such as population growth, technological change, or shifting long-run demand -- for example the decline in death rate due to advances in science, or steadily rising demand for wheat as population grows.

Seasonal variations, S, are short-term movements that recur regularly, with a fixed periodicity of a year or less (daily, weekly, monthly, yearly) -- caused by the calendar seasons, religious festivals, or social customs, such as increased electricity use in summer or an after-Eid sale spike. Cyclical fluctuations, C, are longer-term oscillations about the trend that recur over a period longer than a year but WITHOUT a fixed periodicity (unlike seasonal variations) -- caused by the business cycle, which has four phases: trough (depression, the lowest point), expansion (recovery, the upswing), peak (boom/prosperity, the highest point), and recession (contraction, the downswing). Cyclical fluctuations are less predictable than seasonal variations or trend, making them more consequential for business planning.

Irregular movements, I, are unpredictable, unsystematic, non-recurring changes caused by random events such as wars, floods, earthquakes, strikes, or fires -- also called erratic, accidental, or random variations.

16.3 Analysis of Time Series

Analysing a time series means decomposing it into its separate components for individual study, mainly to support estimation and forecasting. Two standard models combine the four components: the multiplicative model, Y=TxSxCxI, treats trend T in the series' original units while S, C, and I are unitless percentage index numbers; the additive model, Y=T+S+C+I, treats all four components in the series' original units. The multiplicative model is conventionally treated as the standard model for time series analysis.

Coded Time Variable: to simplify trend calculations, the time variable t is coded as x=(t-tbar)/h (or a rescaled version), where tbar=(first period + last period)/2 and h is the constant interval between periods. Since sum(t-tbar)=0 by construction, the coding guarantees sum(x)=0=sum(x^3)=sum(x^5)=..., which is what dramatically simplifies the least squares normal equations used later in the chapter. For an odd number of periods, x=0 is assigned to the exact middle period; for an even number of periods, x=0 falls halfway between the two middle periods (giving half-integer codes unless the unit of measurement is redefined as half a period, which restores whole-number codes).

16.4 Estimation of Secular Trend

Four methods are used to estimate secular trend, in increasing order of objectivity and mathematical rigour. The Method of Free Hand Curve plots the historigram and then draws, by eye, a smooth line or curve that best represents the general movement -- often anchored through the point (mean time, mean y) for extra reliability. It is simple and quick, and smooths out seasonal variations, but is a rough, personally-biased method: different analysts will draw different lines from the same data.

The Method of Semi-Averages divides the observed series into two equal halves (omitting or duplicating the middle value if n is odd), computes the average of each half, and draws a straight line through the two resulting points, each plotted at its half's midpoint in time. Given the two semi-averages y1' (at x1) and y2' (at x2), the trend line's slope is b=(y2'-y1')/(x2-x1) and intercept a=y1'-b.x1; equivalently, for the whole series, b=4(S2-S1)/n^2 where S1, S2 are the sums of the first and second halves and n is the total number of periods. This method is objective and quick, and gives a better approximation than free hand curve (since it rests on a definite mathematical rule), but is heavily affected by extreme values (because it uses arithmetic means) and is only appropriate when the underlying trend is genuinely linear.

The Method of Moving Averages replaces each observation with the average of itself and a fixed number k of neighbouring periods, sliding this window one period at a time across the whole series: a1=(1/k)sum of the first k values, a2=(1/k)sum of values 2 through k+1, and so on; equivalently, using successive k-period totals S1, S2, S3,..., each new total is obtained from the previous one by dropping the oldest value and adding the next one, S2=S1+y_{k+1}-y1, and each average a_i=S_i/k. Each moving average is plotted against the MIDDLE of its k-period window. When k is odd, this middle falls exactly on an observed time period; when k is even, it falls between two periods, requiring a further 2-period 'centring' average of the moving averages themselves to realign them with the actual observed periods (yielding CENTRED moving averages). Moving averages smooth out periodic (seasonal/cyclical) fluctuations most effectively when the averaging period is chosen to match the periodicity of those fluctuations, but the method sacrifices trend values at both ends of the series, is still sensitive to extreme observations, and provides no actual mathematical equation for extrapolating the trend into the future.

The Method of Least Squares fits a definite mathematical curve -- typically a straight line y-hat=a+bx or a parabola y-hat=a+bx+cx^2, with x the coded time value -- by minimizing the sum of squared deviations of observed values from the fitted curve, exactly the same least squares principle developed in Chapter 14 for ordinary regression. For a straight line, the normal equations sum(y)=na+b.sum(x) and sum(xy)=a.sum(x)+b.sum(x^2) give b=[n.sum(xy)-sum(x)sum(y)]/[n.sum(x^2)-(sum x)^2] and a=ybar-b.xbar; when the time variable has been coded so sum(x)=0, these collapse to the much simpler a=ybar=sum(y)/n and b=sum(xy)/sum(x^2). For a parabola, with x coded so sum(x)=0=sum(x^3), the normal equations similarly simplify to give c=[n.sum(x^2 y)-sum(x^2)sum(y)]/[n.sum(x^4)-(sum x^2)^2], a=[sum(y)-c.sum(x^2)]/n, and b=sum(xy)/sum(x^2). The fitted line/curve always passes through the point (xbar, ybar); the residuals sum to zero (sum y = sum y-hat); and the smaller the sum of squared residuals, the better the fit -- which is exactly how a quadratic trend can be objectively confirmed to fit better than a linear one for a given dataset, by comparing their two sum(e^2) values directly.

Shifting the Origin: given an already-fitted trend line y-hat=a+bx (linear) with the origin at some reference time period, moving the origin k periods FORWARD means substituting (x+k) for x, and k periods BACKWARD means substituting (x-k) for x; only the intercept changes (to a+/-bk), never the slope b. For a quadratic trend, the same substitution is made into y-hat=a+bx+cx^2 and the result expanded and regrouped in powers of the new x, changing all three constants a, b, and c.

Important Definitions

What is a time series?: The sequence y1, y2, ..., yn of n observations of a variable Y, recorded according to their time of occurrence t1, t2, ..., tn, usually at equidistant points in time.

What are the signal and the noise in a time series?: The signal, f(t), is the systematic, determined component of variation; the noise, e, is the random, irregular component; symbolically y=f(t)+e.

What is a historigram?: A graph of a time series, with time t on the x-axis and the observed values of Y on the y-axis, successive points joined by line segments -- the first step in examining any time series.

What is secular trend?: A smooth, steady, gradual movement of a time series in the same direction (up or down) over a fairly long period of time, reflecting long-run forces such as population growth or technological change.

What are seasonal variations?: Short-term movements that recur regularly with a fixed periodicity of a year or less, caused by calendar seasons, festivals, or social customs.

What are cyclical fluctuations?: Longer-term oscillations about the trend recurring over more than a year, WITHOUT a fixed periodicity, caused by the business cycle (trough, expansion, peak, recession).

What are irregular movements?: Unpredictable, non-recurring, unsystematic changes caused by random events such as wars, strikes, floods, or fires; also called erratic or accidental variations.

What is the multiplicative model of a time series?: The model Y=TxSxCxI, in which trend T is in the series' original units while S, C, and I are unitless percentage index numbers; conventionally the standard model for time series analysis.

What is the coded time variable, and why is it used?: A rescaled/shifted time variable x=(t-tbar)/h chosen so that sum(x)=0 (and sum(x^3)=0), which greatly simplifies the least squares normal equations used to estimate secular trend.

Why must k-period moving averages be 'centred' when k is even?: Because an even-period moving average falls exactly halfway between two observed time periods; a further 2-period moving average of these averages realigns ('centres') them with the actual observed periods.

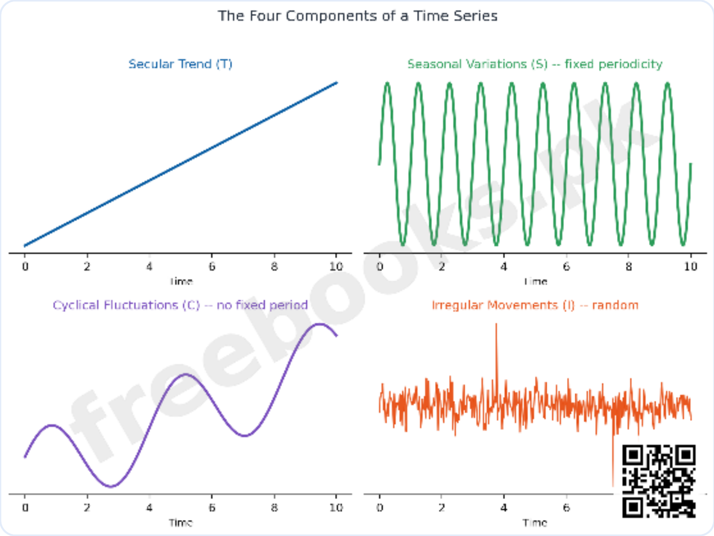
Key Facts and Relations

Topic	Key Fact / Relation
Time series model	y = f(t) + e (f(t) = signal, e = noise)
Multiplicative and additive models	Multiplicative: Y = T x S x C x I; Additive: Y = T + S + C + I

Topic	Key Fact / Relation
Coded time variable (general)	x = (t - tbar)/h, tbar = (first period + last period)/2
Semi-averages trend slope and intercept	b = (y2' - y1')/(x2 - x1); a = y1' - b.x1; [b = 4(S2-S1)/n^2 for whole series]
k-period moving average (successive)	a1 = S1/k; a2 = a1 + (y_{k+1} - y1)/k; S2 = S1 + y_{k+1} - y1
Least squares linear trend (general)	b = [n.sum(xy) - sum(x)sum(y)] / [n.sum(x^2) - (sum x)^2]; a = ybar - b.xbar
Least squares linear trend (coded, sum x=0)	a = ybar = sum(y)/n; b = sum(xy)/sum(x^2)
Quadratic (parabolic) trend model	y-hat = a + bx + cx^2
Quadratic trend estimates (coded, sum x=0=sum x^3)	c = [n.sum(x^2 y) - sum(x^2)sum(y)] / [n.sum(x^4) - (sum x^2)^2]; a = [sum(y) - c.sum(x^2)]/n; b = sum(xy)/sum(x^2)
Shifting the origin (linear trend)	y-hat = a + b(x +/- k) = (a +/- bk) + bx
Sum of squared residuals (goodness of fit)	sum(e^2) = sum(y^2) - a.sum(y) - b.sum(xy) [linear]; minus c.sum(x^2 y) also, for quadratic
Properties of the least squares trend line	Passes through (xbar, ybar); sum(y - y-hat) = 0, so sum(y) = sum(y-hat)

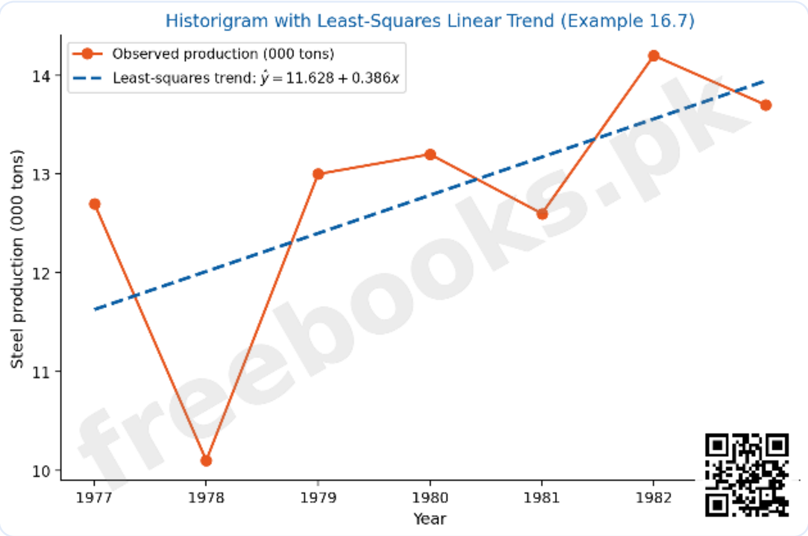
Diagrams

The Four Components of a Time Series: A 4-panel illustration showing idealised patterns for each classical component: secular trend (a smooth rising straight line), seasonal variations (a regular repeating up-down wave pattern within each year), cyclical fluctuations (a longer, irregular-period wave oscillating around a rising trend), and irregular movements (a jagged, unpredictable random pattern with occasional spikes), illustrating how a real time series is the combination of some or all of these



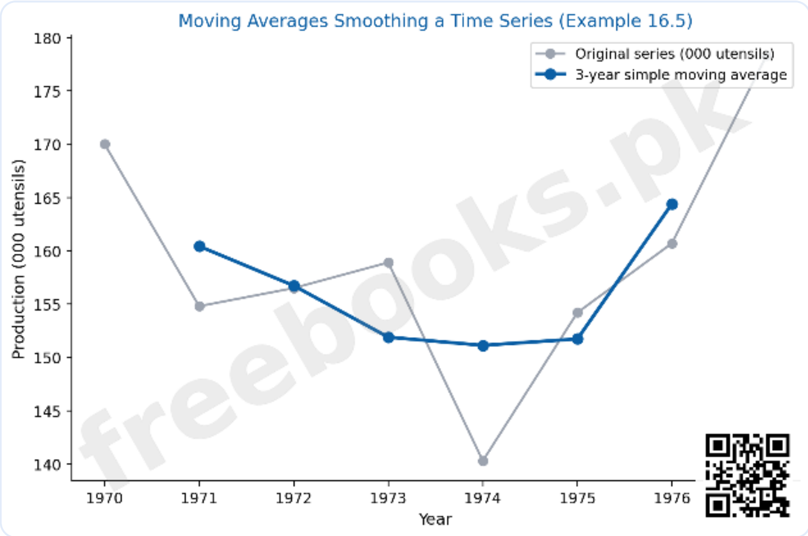
The Four Components of a Time Series

Historigram with Least-Squares Linear Trend (Example 16.7): A historigram (line plot) of steel production 1977-1983, with the least-squares fitted straight trend line y-hat=11.628+0.386x overlaid, showing how the fitted line captures the general upward movement while smoothing over the year-to-year fluctuations



Historigram with Least-Squares Linear Trend (Example 16.7)

Moving Averages Smoothing a Time Series (Example 16.5): The original (jagged) yearly production series for silver utensils 1970-1977 plotted alongside its 3-year simple moving average, showing visually how the moving average smooths out short-term fluctuations to reveal the underlying movement, while losing values at both ends of the series



Moving Averages Smoothing a Time Series (Example 16.5)

Short Questions & Answers

Define a time series and give two examples.

A time series is the sequence of observations of a variable recorded according to their time of occurrence, usually at equidistant points; examples include the monthly sales of a store and the annual wheat yield of a country.

Distinguish between seasonal variations and cyclical fluctuations.

Seasonal variations recur with a FIXED periodicity of a year or less (daily/weekly/monthly/yearly) caused by calendar seasons or customs; cyclical fluctuations recur over MORE than a year WITHOUT a fixed periodicity, caused by the business cycle, and are less predictable.

Why is the time variable coded before estimating secular trend?

Coding shifts and rescales the origin so that sum(x)=0 (and sum(x^3)=0 for quadratic trend), which greatly simplifies the least squares normal equations, reducing them to simple formulas like a=ybar and b=sum(xy)/sum(x^2).

State one merit and one demerit of the method of moving averages.

Merit: an appropriately chosen period eliminates seasonal/cyclical fluctuations effectively. Demerit: it does not provide trend values at the beginning and end of the series, and gives no mathematical equation for forecasting.

How is a k-period moving average 'centred' when k is even?

By computing a further 2-period moving average of the original k-period moving averages, which realigns them to correspond with the actual observed time periods rather than falling between two of them.

What effect does shifting the origin of a linear trend equation have on its slope and intercept?

The slope b never changes; only the intercept changes, becoming a+bk (shifting k periods forward) or a-bk (shifting k periods backward).

Long Questions & Answers

Describe the four classical components of a time series in detail, and explain how the multiplicative and additive models combine them, illustrating with realistic examples for each component.

A time series records the same variable repeatedly over successive points in time, and a central insight of this chapter is that the resulting sequence of values is rarely a single, uniform kind of movement -- instead, it is generally the combined outcome of up to four distinct, separately-caused kinds of variation acting together, and separating these out is the whole purpose of time series analysis. The secular (or long-term) trend, denoted T, captures the smooth, gradual, persistent movement of the series in one consistent direction -- either steadily rising or steadily falling -- over a fairly long span of time, reflecting slow-moving underlying forces that don't reverse direction from one period to the next. Classic examples include the long-run decline in death rates as medical science and public health advance, the steadily increasing demand for wheat as a country's population grows year after year, or the gradual, sustained shift in automobile demand toward smaller, more fuel-efficient models as fuel costs and environmental awareness rise over decades. Seasonal variations, denoted S, are an entirely different kind of movement: short-term fluctuations that repeat with a FIXED, predictable periodicity of one year or less -- daily, weekly, monthly, or annually -- driven by causes tied directly to the calendar itself, such as the physical seasons, recurring religious festivals, or established social customs. A department store's weekly sales pattern (busier on weekends), a sharp jump in electricity consumption every summer as air conditioning use rises, or the predictable post-Eid clearance sale spike are all textbook seasonal variations: the KEY diagnostic feature is that both the timing and the approximate size of the fluctuation repeat reliably, period after period, in a way that can be anticipated in advance. Cyclical fluctuations, denoted C, superficially resemble seasonal variations in that they too are recurring, wave-like oscillations around the underlying trend line, but they differ in one crucial respect: they do NOT have a fixed periodicity. A business cycle -- the classic source of cyclical fluctuation in economic time series -- might last two years in one instance and seven years in the next, moving irregularly through four recognised phases: a trough (or depression), the lowest point of the cycle; an expansion (or recovery), the upswing phase as the trough phase reverses and activity begins climbing; a peak (or boom/prosperity), the highest point the cycle reaches; and finally a recession (or contraction), where the upswing turns and begins declining again toward the next trough. Because cyclical fluctuations lack the seasonal component's fixed, predictable rhythm, they are inherently harder to forecast and, correspondingly, more dangerous for business and economic planning than seasonal variations, even though both are technically 'recurring' in character. Finally, irregular (or erratic, accidental, random) movements, denoted I, capture whatever variation is left over once trend, seasonal, and cyclical effects have all been accounted for -- unpredictable, non-recurring, one-off disturbances caused by genuinely random external events such as wars, natural disasters like floods or earthquakes, labour strikes, factory fires, or sudden political upheavals; a steel strike that delays production for a single week, or a factory fire that halts output for three weeks, are exactly this kind of component, precisely because neither event follows any discernible pattern that could ever be anticipated or built into a forecasting model. Once all four of these conceptually distinct components have been identified, the analyst must decide how they combine together to produce the ACTUAL observed value Y at any given moment, and two standard mathematical models are used for this purpose. The multiplicative model expresses this combination as a straightforward product, $Y=T\times S\times C\times I$, under which the trend component T is expressed in the series' own original units (e.g. actual tons of steel, or actual rupees of sales), while the remaining three components S, C, and I are all expressed as unitless PERCENTAGE index numbers that scale the trend value up or down (a seasonal index of 120% for December sales would indicate December sales run 20% above what the trend alone would predict, for instance). By long-standing convention, this multiplicative model is treated as the STANDARD model for time series analysis throughout applied statistics. The alternative additive model instead expresses the combination as a simple sum, Y=T+S+C+I, under which ALL FOUR components -- not just the trend -- must be expressed consistently in the series' own original units (e.g. all measured in actual rupees, or actual units sold, rather than mixing raw units for trend with percentage indices for the other three); this model implicitly assumes that the absolute SIZE of the seasonal, cyclical, and irregular fluctuations stays roughly constant over time, regardless of how large or small the underlying trend level happens to be at that particular moment, which is often a less realistic assumption for economic and business series (where fluctuations more commonly scale up proportionally as the trend level itself grows larger) -- helping to explain why the multiplicative model has become the conventional default choice in practice.

Compare the four methods of estimating secular trend -- free hand curve, semi-averages, moving averages, and least squares -- explaining the mechanics, and the relative merits and demerits, of each, and explain why least squares is generally considered the most rigorous of the four.

Estimating the secular trend of a time series -- isolating the smooth, long-run directional movement from the noisier seasonal, cyclical, and irregular fluctuations superimposed on top of it -- can be approached through four progressively more rigorous and objective methods, and understanding the trade-offs between them is essential to choosing the right one for a given practical situation. The method of free hand curve is the simplest and most intuitive: after plotting the historigram of the observed series, the analyst simply draws, purely by eye, a smooth straight line or curve through the scatter of points that seems to best capture the general upward or downward movement, ideally anchored through the point representing the overall mean of the time period and the overall mean of the observed values, since this is a reasonable condition the true trend line should satisfy. This method is fast, requires no calculation whatsoever, and effectively smooths out the noisier seasonal fluctuations by eye -- but its fundamental weakness is that it is entirely subjective: given the identical dataset, two different analysts, or even the same analyst on two different days, could easily draw two visibly different trend lines, and there is no principled way to adjudicate which one is objectively 'more correct,' making the method unreliable for any application requiring reproducibility or defensible precision. The method of semi-averages introduces a first, modest dose of mathematical objectivity: the observed series is split into exactly two equal halves (with the single middle observation either dropped entirely, or duplicated into both halves, if the total number of periods happens to be odd), the arithmetic mean of each half is computed and plotted at that half's own midpoint in time, and a single straight line is then drawn connecting these two computed points, with its slope and intercept determined by ordinary two-point coordinate geometry rather than by eye. This produces a genuinely objective result, in the specific sense that any two analysts working from the identical data will always arrive at EXACTLY the same trend line -- a clear improvement over the free hand method -- but it carries its own serious limitations: because it relies on simple arithmetic means, it remains highly sensitive to unusually extreme values occurring within either half, and more fundamentally, the method is only appropriate at all when the underlying trend is genuinely linear (or very close to it); it offers no mechanism whatsoever for handling a visibly curved, non-linear trend pattern. The method of moving averages takes a rather different approach: instead of attempting to fit any single overall curve to the whole series at once, it works locally, replacing each individual observation with the average of itself and a symmetric window of k neighbouring periods, sliding this averaging window one period forward at a time across the entire length of the series (dropping the oldest observation from the window and including the next new one each time it slides). Because averaging over several adjacent periods tends to cancel out short-term seasonal and irregular noise while a genuine underlying trend movement tends to persist and reinforce itself across the averaging window, this method can be remarkably effective at revealing the trend -- PROVIDED the chosen averaging period k is deliberately matched to the actual periodicity of whatever seasonal or cyclical fluctuation needs to be eliminated (for instance, using a 12-month moving average to eliminate a fluctuation with an annual, 12-month period). Its notable drawbacks, however, are that no trend values can be computed at all near the very beginning and very end of the series (since a full symmetric window of k periods simply isn't available there), the resulting moving-average values remain sensitive to any individual extreme observations that happen to fall within a given averaging window, and critically, the method never produces any actual mathematical EQUATION for the trend -- meaning it cannot, by itself, be used to forecast or extrapolate values into future time periods beyond the observed data, unlike the other three methods. The method of least squares is generally regarded as the most rigorous and, correspondingly, the most widely used of the four available methods, precisely because it fits a definite, explicit mathematical equation -- most commonly a straight line y-hat=a+bx, though a quadratic (parabolic) curve y-hat=a+bx+cx^2 is used instead whenever a straight line visibly fails to capture genuine curvature present in the data -- chosen specifically so as to make the sum of the SQUARED deviations between every observed value and its corresponding fitted trend value as small as mathematically possible, following exactly the identical least squares principle already developed for ordinary two-variable regression back in Chapter 14, merely with coded time itself now standing in as the independent variable x. Because this method yields a definite, fully-specified equation rather than merely a hand-drawn line or a table of local averages, it can be used both to compute a precise trend value at any observed historical time period AND to extrapolate forward, generating genuine numerical forecasts for future time periods purely by substituting new, larger values of x into the fitted equation. Additionally, because different candidate models (for instance, a straight line versus a quadratic curve, fitted to that exact same underlying dataset) can be directly and objectively compared against one another simply by comparing their two respective sums of squared residuals side by side -- the model producing the smaller sum of squared residuals is, by the very definition of the least squares criterion itself, unambiguously the better-fitting model for that particular dataset -- the least squares method offers a level of built-in, principled rigour that none of the three preceding methods can genuinely match, which is precisely why it has become the standard, default method of choice whenever a reasonably precise, forecastable trend equation is genuinely required in serious practical or academic work.

Multiple Choice Questions (MCQs)

A time series is best described as: (A) Any collection of numerical data (B) A sequence of observations recorded according to their time of occurrence (C) Only annual data (D) Data with no order at all

Correct answer: (B) A sequence of observations recorded according to their time of occurrence. A time series is specifically the sequence y1,...,yn recorded according to time of occurrence t1,...,tn -- order in time is essential.

A graph of a time series, with time on the x-axis, is called a: (A) Histogram (B) Scatter diagram (C) Historigram (D) Ogive

Correct answer: (C) Historigram. A historigram plots the observed values of Y against time t, joined by line segments -- distinct from a histogram (a bar chart of frequency distributions).

Which time series component recurs with a FIXED periodicity of one year or less? (A) Secular trend (B) Seasonal variations (C) Cyclical fluctuations (D) Irregular movements

Correct answer: (B) Seasonal variations. Seasonal variations have a fixed periodicity (daily/weekly/monthly/yearly); cyclical fluctuations recur but WITHOUT a fixed periodicity.

The four phases of a business cycle, in order, are: (A) Peak, trough, recession, expansion (B) Trough, expansion, peak, recession (C) Expansion, peak, trough, recession (D) Recession, trough, peak, expansion

Correct answer: (B) Trough, expansion, peak, recession. The standard order is: trough (depression) -> expansion (recovery) -> peak (boom) -> recession (contraction) -> back to trough.

In the multiplicative model Y = T x S x C x I, the components S, C, and I are expressed as: (A) Original units of Y, like T (B) Unitless percentage index numbers (C) Always equal to 1 (D) Negative numbers

Correct answer: (B) Unitless percentage index numbers. In the multiplicative model, only T is in Y's original units; S, C, and I are unitless percentage index numbers.

Coding the time variable so that sum(x) = 0 is done primarily to: (A) Make the data more accurate (B) Simplify the least squares normal equations (C) Eliminate seasonal variation (D) Remove irregular movements

Correct answer: (B) Simplify the least squares normal equations. sum(x)=0 (and sum(x^3)=0) collapses the least squares normal equations to simple formulas like a=ybar, b=sum(xy)/sum(x^2).

The main weakness of the free hand curve method for estimating trend is that it is: (A) Too mathematically complex (B) Subject to personal bias (C) Unable to handle any data at all (D) Only usable for seasonal data

Correct answer: (B) Subject to personal bias. Free hand curve fitting is drawn by eye, so different analysts can draw different trend lines from the same data -- it is highly subjective.

A k-period moving average needs to be 'centred' when: (A) k is odd (B) k is even (C) The trend is quadratic (D) n is less than 10

Correct answer: (B) k is even. When k is even, the moving average falls between two observed periods; a further 2-period average 'centres' it back onto an observed period.

If a linear trend equation is y-hat = a + bx with origin at year 1980, shifting the origin FORWARD by k years gives a new intercept of: (A) a - bk (B) a + bk (C) a x bk (D) a / bk

Correct answer: (B) a + bk. Shifting the origin forward by k substitutes (x+k) for x, giving y-hat=a+b(x+k)=(a+bk)+bx -- the new intercept is a+bk.

When comparing a linear trend and a quadratic trend fitted to the same data, the better-fitting model is the one with: (A) The larger slope b (B) The smaller sum of squared residuals, sum(e^2) (C) The larger intercept a (D) More data points

Correct answer: (B) The smaller sum of squared residuals, sum(e^2). By the least squares criterion, the model with the smaller sum(e^2) fits the observed data more closely and is preferred.

Quick Revision Summary

- Time series: y=f(t)+e, signal f(t) + noise e; historigram = time-series line graph
- 4 components: Secular trend (T, long-term smooth) | Seasonal (S, fixed periodicity <=1yr) | Cyclical (C, no fixed periodicity, business cycle) | Irregular (I, random/unpredictable)
- Business cycle phases: trough -> expansion -> peak -> recession
- Multiplicative: Y=TxSxCxI (T in original units, S/C/I unitless %) | Additive: Y=T+S+C+I (all original units)
- Coded time variable: x=(t-tbar)/h chosen so sum(x)=0=sum(x^3), simplifying least squares
- Free hand curve: quick but subjective/biased | Semi-averages: objective but only for linear trend, sensitive to extremes
- Moving averages: smooth out fixed-period fluctuations; no trend values at ends; no forecasting equation
- Least squares linear: b=sum(xy)/sum(x^2), a=ybar (when sum x=0) | Quadratic: adds c=[n.sum(x^2y)-sum(x^2)sum(y)]/[n.sum(x^4)-(sum x^2)^2]
- Best-fit model = smaller sum(e^2) = sum(y^2)-a.sum(y)-b.sum(xy)[-c.sum(x^2y)]
- Shifting origin (linear): only intercept changes, a+/-bk; slope b unchanged

Exam Tips

- Always draw the historigram first -- visual inspection tells you whether to expect a linear or curved trend before you compute anything
- Remember: seasonal = FIXED periodicity, cyclical = NO fixed periodicity -- this is the single most-tested distinction in this chapter
- When n is even and using semi-averages, the two midpoint x-values (x1, x2) will be at the centre of each half, not at integer years -- don't forget this when computing b=(y2'-y1')/(x2-x1)
- For moving averages, always double-check whether k is odd (no centring needed) or even (must centre with a further 2-period average)

- When coding time for least squares, decide up front whether the unit is 1 year or 1/2 year for EVEN numbers of periods -- this changes b (but not the trend values themselves) by a factor of 2
- To compare linear vs quadratic fits, compute BOTH sum(e^2) values and pick the smaller one -- don't just eyeball which curve 'looks' better