# DEMAND ESTIMATION AND FORECASTING

Small Business Development (EP60029)

# Empirical Demand Functions

- Demand equations derived from actual market data
- Useful in making pricing & production decisions
- Representative sample: a sample usually drawn randomly, that has characteristics that accurately reflect the population as a whole
- Response bias: the difference between the response given by an individual to a hypothetical question and the action the individual takes when the situation actually occurs

# Linear Empirical Demand Specification

- In *linear* form, an empirical demand function can be specified:

$$Q = a + bP + cM + dP_R + eN$$

where, Q is quantity demanded, P is the price of good/service, M is consumers' income, $P_R$ is the price of related good R, & N is the number of buyers

- In above linear form
  - ***b = ΔQ/ΔP, assumed to be negative***
  - ***c = ΔQ/ΔM >0, normal good***
  - ***c = ΔQ/ΔM <0, inferior good***
  - ***d = ΔQ/ΔP$_R$ >0, substitute good***
  - ***d = ΔQ/ΔP$_R$ >0, complement good***
  - ***d = ΔQ/ΔN, assumed to be positive***

# Nonlinear Empirical Demand Functions

- Most common nonlinear demand specification is log-linear (or constant elasticity)

$$Q = aP^{b}M^{c}P_{R}^{d}N^{e}$$

- The linear function to be estimated is converted using natural logarithms

$$\ln Q = \ln a + b \ln P + c \ln M + d \ln P_R + e \ln N$$

# DEMAND: Basic Estimation Techniques

# Simple Linear Regression

- **Simple linear regression model relates dependent variable $Y$ to one independent (or explanatory) variable $X$**

$$Y = a + bX$$

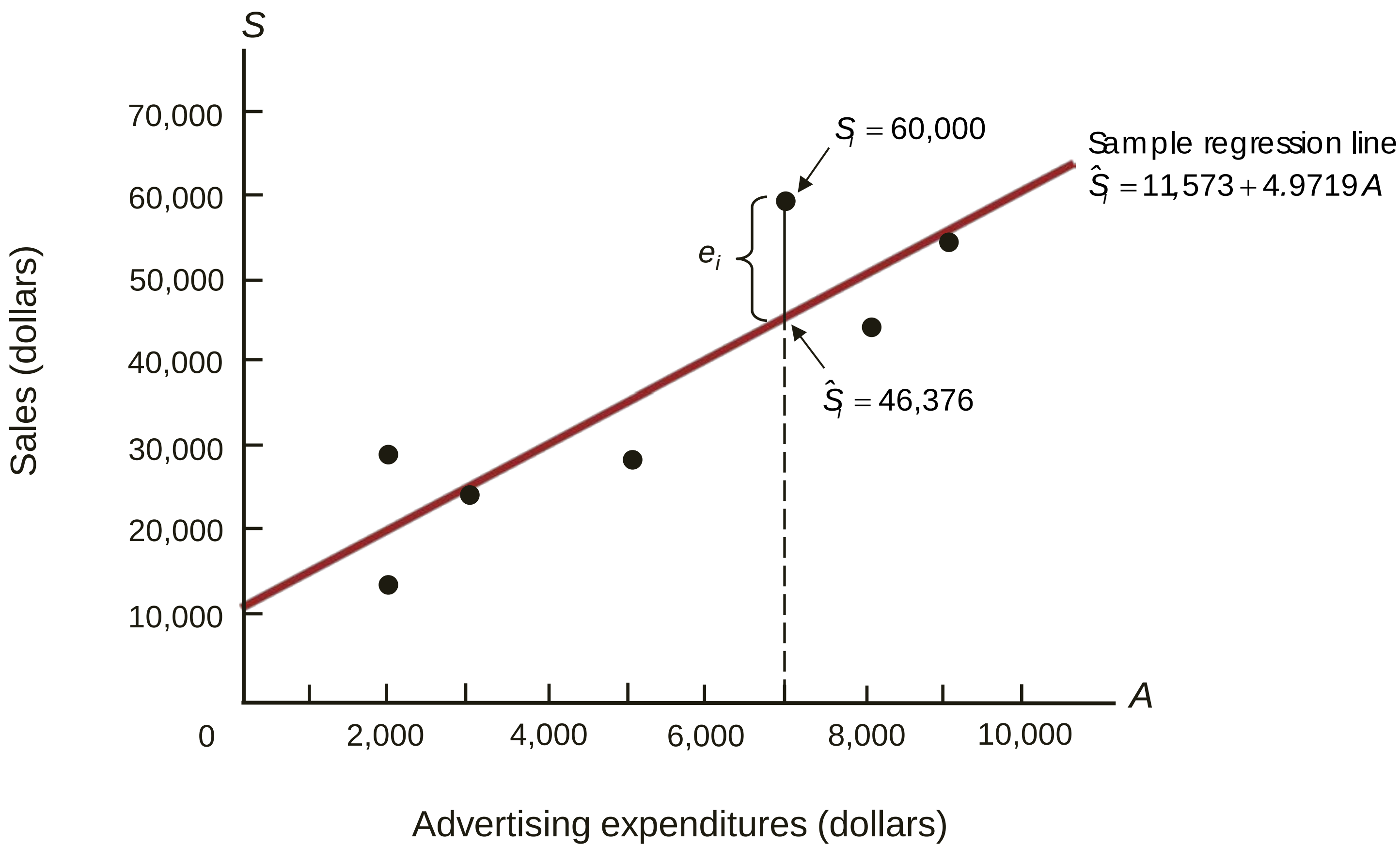
  - Intercept parameter ($a$) gives value of $Y$ where regression line crosses $Y$-axis (value of $Y$ when $X$ is zero)
  - Slope parameter ($b$) gives the change in $Y$ associated with a one-unit change in $X$, $b = \Delta Y / \Delta X$

# Method of Least Squares

- **Parameter estimates are obtained by choosing values of $a$ & $b$ that minimize the sum of squared residuals**
  - The residual is the difference between the actual & fitted values of $Y$, $Y_i - \hat{Y}_i$
- **The sample regression line is an estimate of the true regression line**

$$\hat{Y} = \hat{a} + \hat{b}X$$

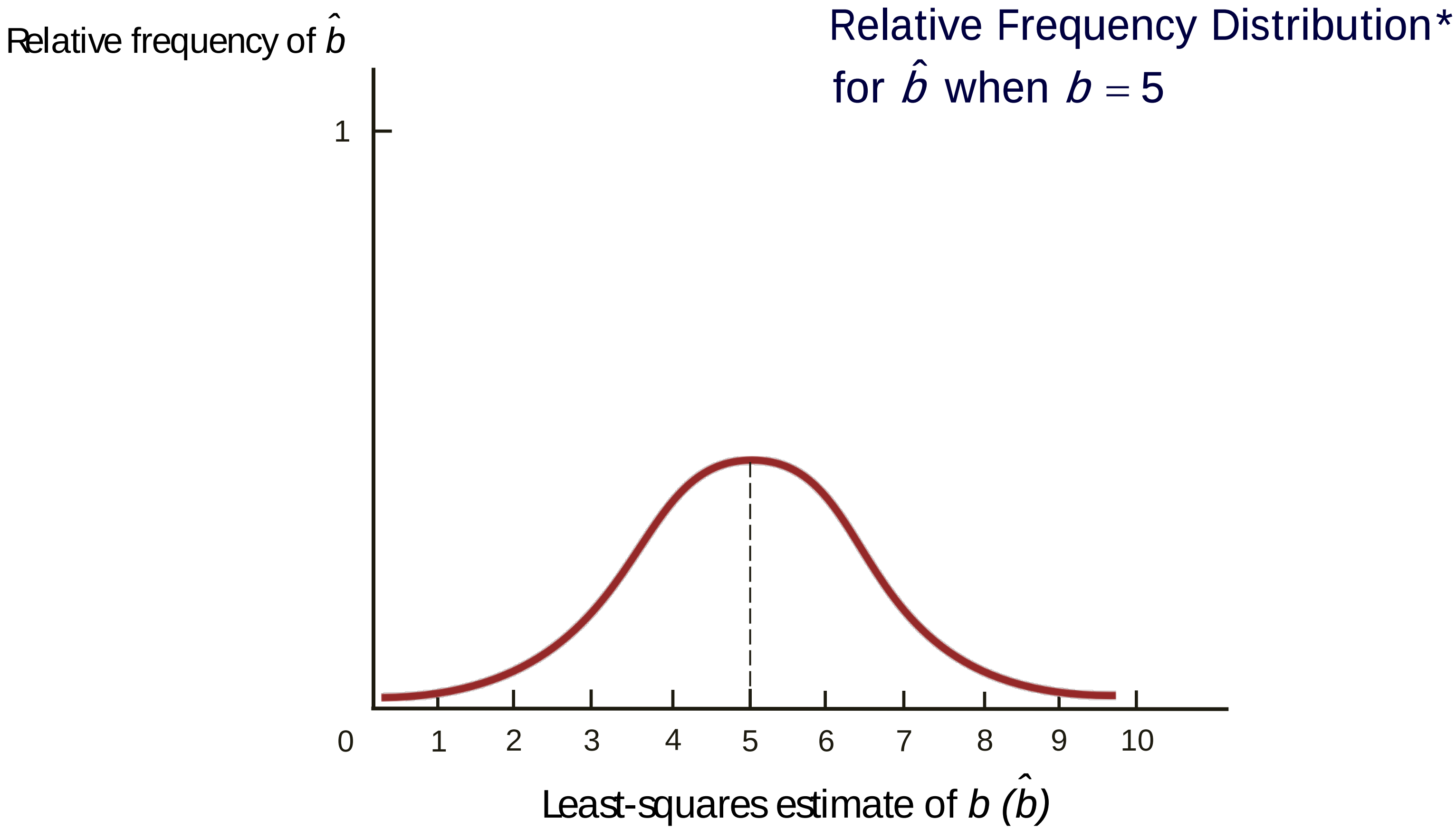
# Sample Regression Line



# Unbiased Estimators

- **The estimates of $\hat{a}$ & $\hat{b}$ do not generally equal the true values of $a$ & $b$**
  - $\hat{a}$ & $\hat{b}$ are random variables computed using data from a random sample
- The distribution of values the estimates might take is centered around the true value of the parameter
- An estimator is *unbiased* if its average value (or expected value) is equal to the true value of the parameter

# Relative Frequency Distribution*




*Also called a probability density function (pdf)

# Statistical Significance

- Must determine if there is sufficient statistical evidence to indicate that $Y$ is truly related to $X$ (i.e., $b \neq 0$)

- **Even if $b = 0$ it is possible that the sample will produce an estimate $\hat{b}$ that is different from zero**
- **Test for statistical significance using $t$-tests or $p$-values**

# Performing a t-Test

- First determine the *level of significance*
  - Probability of finding a parameter estimate to be statistically different from zero when, in fact, it is zero
  - Probability of a *Type I Error*
- 1 – *level of significance* = *level of confidence*

# Performing a t-Test

- $t$-ratio is computed as $t = \dfrac{\hat{b}}{S_{\hat{b}}}$

  where $S_{\hat{b}}$ is the standard error of the estimate $\hat{b}$

- Use $t$-table to choose critical $t$-value with $n - k$ degrees of freedom for the chosen level of significance
  - $n =$ number of observations
  - $k =$ number of parameters estimated

# Performing a t-Test

- If absolute value of $t$-ratio is greater than the critical $t$, the parameter estimate is statistically significant

# Using p-Values

- Treat as statistically significant only those parameter estimates with $p$-values smaller than the maximum acceptable significance level
- $p$-value gives exact level of significance
  - Also the probability of finding significance when none exists

# Coefficient of Determination

- $R^2$ measures the percentage of total variation in the dependent variable that is explained by the regression equation
  - Ranges from 0 to 1
  - High $\boldsymbol{R^2}$ indicates $\boldsymbol{Y}$ and $\boldsymbol{X}$ are highly correlated

# F-Test

- Used to test for significance of overall regression equation
- Compare $F$-statistic to critical $F$-value from $F$-table
  - Two degrees of freedom, $n - k$ & $k - 1$
  - Level of significance
- If $F$-statistic exceeds the critical $F$, the regression equation overall is statistically significant

# Multiple Regression

- Uses more than one explanatory variable
- Coefficient for each explanatory variable measures the change in the dependent variable associated with a one-unit change in that explanatory variable

# Quadratic Regression Models

- Use when curve fitting scatter plot

**is ∪-shaped or ∩-shaped**

- $Y = a + bX + cX^2$
  - For linear transformation compute new variable $Z = X^2$
  - Estimate $Y = a + bX + cZ$

# Log-Linear Regression Models

- **Use when relation takes the form:** $Y = aX^{b}Z^{c}$

- $b = \dfrac{\text{Percentage change in } Y}{\text{Percentage change in } X}$

- $c = \dfrac{\text{Percentage change in } Y}{\text{Percentage change in } Z}$

- Transform by taking natural logarithms:
  $\ln Y = \ln a + b \ln X + c \ln Z$

- $b$ and $c$ are elasticities

# Market Determined vs. Manager Determined Prices

- Method of estimating parameters of an empirical demand function depends on whether price of the product is *market-determined* or *manager-determined*
- Price-taking firms do not set the price of their product
  - Prices are *endogenous*, or *market-determined* by the intersection of demand & supply
- For price-setting firms
  - Prices are *manager-determined*, or *exogenous*

# Industry Demand for a Price-taker

- To estimate industry demand function for a price-taking firm:
  - ***Step 1:*** Specify industry demand & supply equations
  - ***Step 2:*** Check for identification of industry demand
  - ***Step 3:*** Collect data for the variables in demand and supply
  - ***Step 4:*** Estimate industry demand

# Estimating Demand for a Price-setting Firm

- ***Step 1:*** *Specify the price-setting firm's demand function*
- ***Step 2:*** Collect data on the variable in the firm's demand function
- ***Step 3:*** *Estimate the price-setting firm's demand*

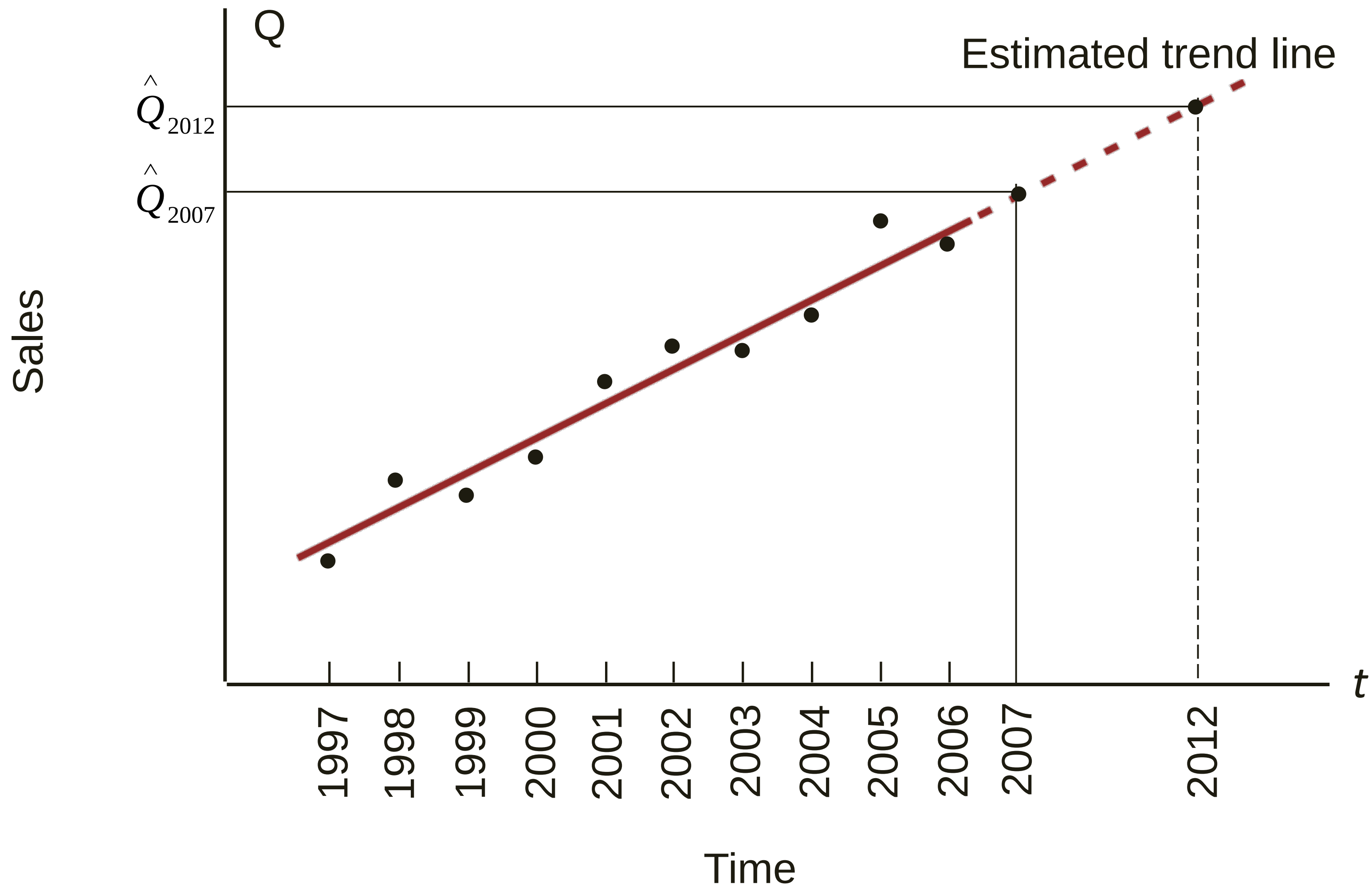
# Time Series Forecasts

- Time-series is a statistical model that shows how a time-ordered sequence of observations on a variable is generated

- Simplest form is linear trend forecasting
  - Sales in each time period ***(Q**_t_)* are assumed to be linearly related to time *(**t**)*

$$Q_t = a + b_t$$

# Linear Trend Forecasting



# Linear Trend Forecasting

- Linear relation between Sales and Time

$$\hat{Q}_t = a + bt$$

- The estimated trend line

$$\hat{Q}_t = \hat{a} + \hat{b}\,t$$

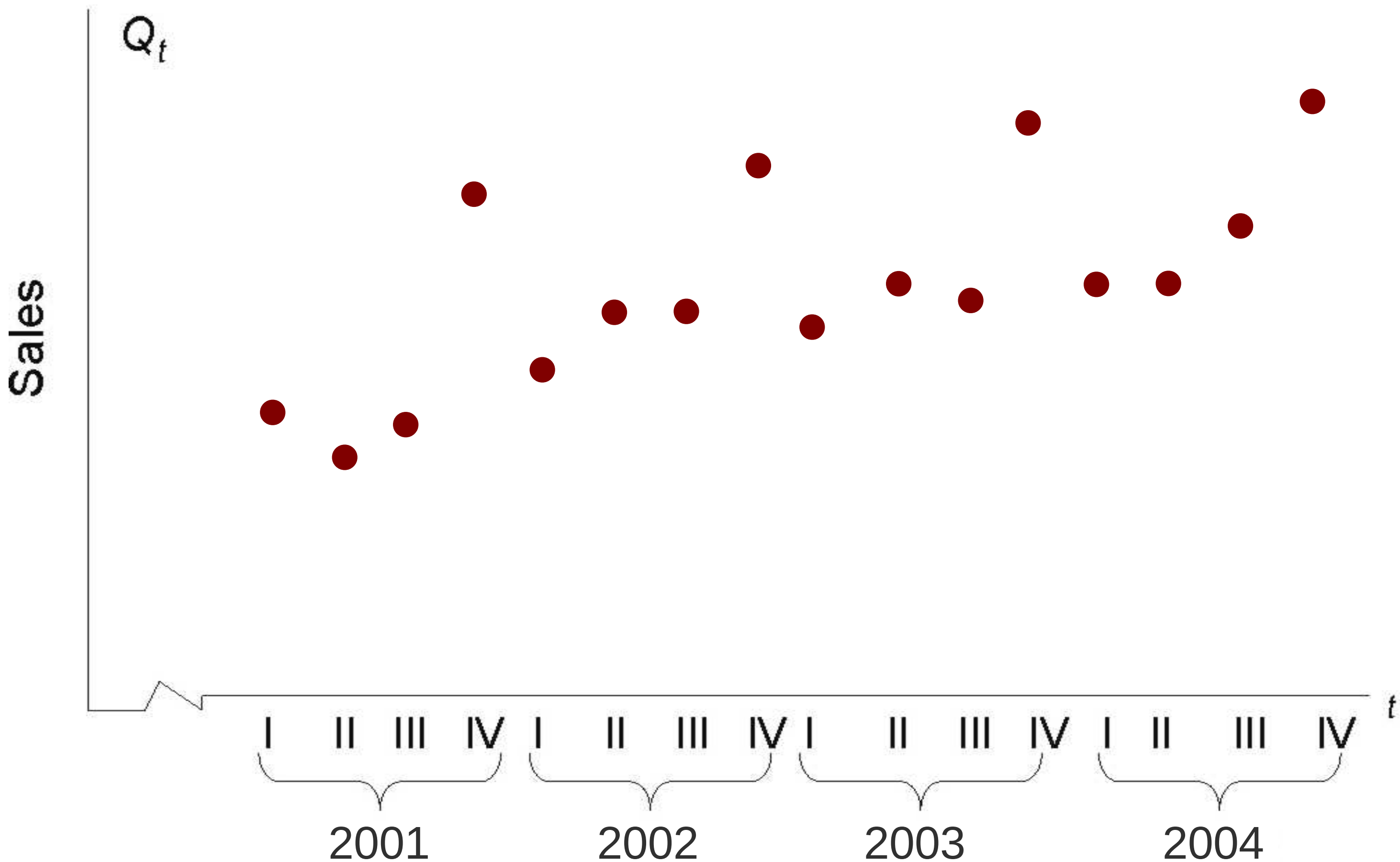
- Forecast for Sales in specific year

$$\hat{Q}_{2007} = a + b \times (2007)$$

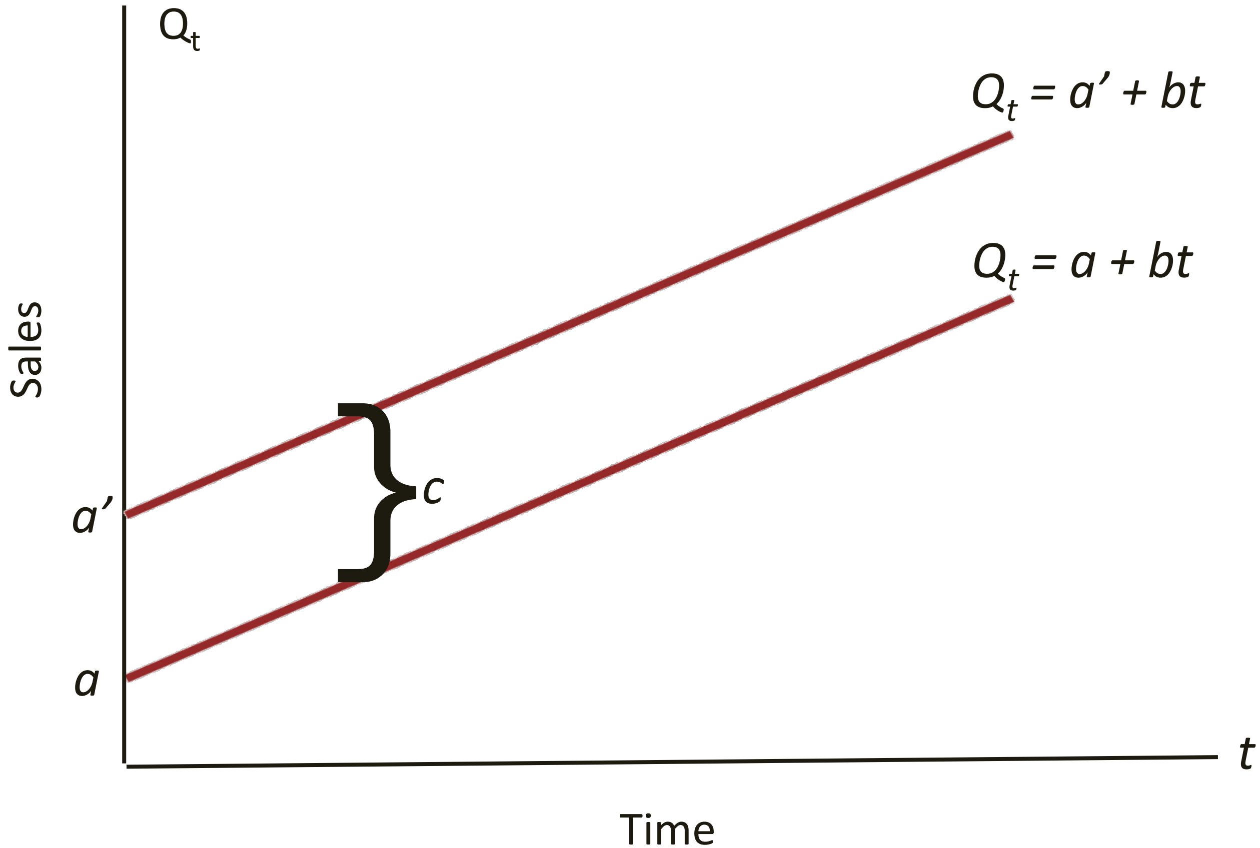
# Seasonal (or Cyclical) Variations

- Regular variation that time-series data exhibit frequently
- Can bias the estimation of parameters in linear trend forecasting
- To account for such variation, *dummy variables* are added to the trend equation
  - Shift trend line up or down depending on the particular seasonal pattern
  - Significance of seasonal behavior determined by using ***t***-test or ***p***-value for the estimated coefficient on the dummy variable

# Sales with Seasonal Variation



# Effect of Seasonal Variation



# Creating a Dummy Variable

- $Q_t = a + bt + cD$
- For quarters I, II, and III the estimated intercept is $\hat{a}$
- For the IVth quarter the estimated intercept is $\hat{a} + \hat{c}$
- For any future period t (=I, II, & III), the sales forecast would be, $\hat{Q}_t = \hat{a} + \hat{b}t$
- For period IV, $\hat{Q}_t = \hat{a} + \hat{b}t + \hat{c} = \left(\hat{a} + \hat{c}\right) + \hat{b}t$
- If there exists quarter-to-quarter differences in sales, then equation is

$$Q_t = a + bt + c_1 D_1 + c_2 D_2 + c_3 D_3$$

| $Q_t$ | t | D |
|---|---|---|
| $Q_{2004(I)}$ | 1 | 0 |
| $Q_{2004(II)}$ | 2 | 0 |
| $Q_{2004(III)}$ | 3 | 0 |
| $Q_{2004(IV)}$ | 4 | 1 |
| $Q_{2005(I)}$ | 5 | 0 |
| $Q_{2005(II)}$ | 6 | 0 |
| $Q_{2005(III)}$ | 7 | 0 |
| $Q_{2005(IV)}$ | 8 | 1 |
| $Q_{2006(I)}$ | 9 | 0 |
| $Q_{2006(II)}$ | 10 | 0 |
| $Q_{2006(III)}$ | 11 | 0 |
| $Q_{2006(IV)}$ | 12 | 1 |
| $Q_{2007(I)}$ | 13 | 0 |
| $Q_{2007(II)}$ | 14 | 0 |
| $Q_{2007(III)}$ | 15 | 0 |
| $Q_{2007(IV)}$ | 16 | 1 |

# Dummy Variables

- A variable that takes only values of 0 and 1
- To account for *N* seasonal time periods
  - ***N – 1*** dummy variables are added
- Each dummy variable accounts for one seasonal time period
  - Takes value of 1 for observations that occur during the season assigned to that dummy variable
  - Takes value of 0 otherwise

**Example: Jean Reynolds, the sales manager of Statewide Trucking Co. wishes to predict sales for all four quarter of 2008. The sales are subject to seasonal variation & also have a trend over time. Reynolds obtain sales data for 2004-07 by quarter. Jean knows that obtaining the desired sales forecast requires to estimate an equation containing three dummy variables – one less than the no. of time periods in the annual cycle. Graphically show the trend and forecast sales for all the four quarters for 2008.**

| Year | Quarter | Sales ($) | t | D1 | D2 | D3 |
|---|---|---|---|---|---|---|
| 2004 | I | 72, 000 | 1 | 1 | 0 | 0 |
| | II | 87, 000 | 2 | 0 | 1 | 0 |
| | III | 87, 000 | 3 | 0 | 0 | 1 |
| | IV | 150, 000 | 4 | 0 | 0 | 0 |
| 2005 | I | 82, 000 | 5 | 1 | 0 | 0 |
| | II | 98, 000 | 6 | 0 | 1 | 0 |
| | III | 94, 000 | 7 | 0 | 0 | 1 |
| | IV | 162, 000 | 8 | 0 | 0 | 0 |
| 2006 | I | 97, 000 | 9 | 1 | 0 | 0 |
| | II | 105, 000 | 10 | 0 | 1 | 0 |
| | III | 109, 000 | 11 | 0 | 0 | 1 |
| | IV | 176, 000 | 12 | 0 | 0 | 0 |
| 2007 | I | 105, 000 | 13 | 1 | 0 | 0 |
| | II | 121, 000 | 14 | 0 | 1 | 0 |
| | III | 119, 000 | 15 | 0 | 0 | 1 |
| | IV | 180, 000 | 16 | 0 | 0 | 0 |

| Dependent Variable: QT | | R-square | t ratio | p-value on F |
|---|---|---|---|---|
| Observations: 16 | | 0.9965 | 794.126 | 0.0001 |
| Variable | Parameter | Standard | | |
| | Estimate | Error | t - ratio | p - value |
| Intercept | 139625.0 | 1743.6 | 80.08 | 0.0001 |
| T | 2737.5 | 129.96 | 21.06 | 0.0001 |
| D1 | - 69788.0 | 1689.5 | - 41.31 | 0.0001 |
| D2 | - 58775.0 | 1664.3 | - 35.32 | 0.0001 |
| D3 | - 62013.0 | 1649.0 | - 37.61 | 0.0001 |

# Trend Projection

$$S_t = S_0 + bt$$

- $S_t$ is the value of the time series for period t, $S_0$ is the estimated value of the time series in the base period (at t=0), b is the absolute amount of growth per period and t is the time period in which the time series is to be forecast. E.g. $S_t = 11.90 + 0.394t$ substituting the value of t, we get the forecast value in period t.
- In many cases the constant percentage growth rate model is appropriate:
- $S_t = S_0(1+g)^t$ , to estimate g first transform time series data into their natural logarithm in linear form. Hence, the equation will be as follows:
- $\ln S_t = \ln S_0 + t\ln(1+g)$ where value of $\ln S_0$ is its antilog (e.g. antilog of $\ln S_0 = 2.49$ is $S_0 = 12.06$ and antilog of $\ln(1 + g) = 0.026$ gives $(1 + g) = 1.026$). Substituting these values back into equation, we get $S_t = 12.06(1.026)^t$ where $S_0 = 12.06$ (at t=0) and estimated growth rate is 1.026 (2.6%)
- For future estimates the value of t is substituted.

# Smoothing Techniques

- **Moving Averages**
  - Forecast value of a time series in a given period is equal to the average value of the time series in a no. of previous periods
  - E.g. with a 3 time period MA, forecast value for next time period is given by avg. value of time series in previous 3 periods
  - More useful when more erratic/random time series data
  - To decide which MA forecast is better we calculate **root mean square error (RMSE)** where $A_t$ is actual value of time series in period t, $F_t$ is forecast value and n is no. of time periods / obs.
  - $$RMSE = \sqrt{\frac{\sum (A_t - F_t)^2}{n}}$$ On comparing RMSE value, the lower the error value, the better is forecast

# Smoothing Techniques

- **Exponential Smoothing**
  - Forecast for period t+1 (i.e., $F_{t+1}$) is a weighted avg. of the actual and forecast values of the time series in period t
  - The value of time series at period t ($A_t$) is assigned a wt. (w) between 0 and 1 inclusive, and the forecast for period t (i.e., $F_t$) is assigned the wt. of 1-w
  - The greater the value of w the greater is the wt. given to the value of time series in period t as opposed to previous periods. Thus:

$$F_{t+1} = wA_t + (1 - w)F_t$$

  - Different values of w are tried, the one leading to smallest RMSE is actually used in forecasting

# Some Final Warnings

- Model misspecification, either by excluding an important variable or by using an inappropriate functional form, reduces reliability of the forecast
- Forecasts are incapable of predicting sharp changes that occur because of structural changes in the market

# THE END