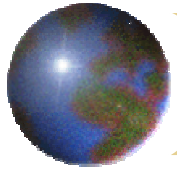


# **Demand Estimation & Forecasting**

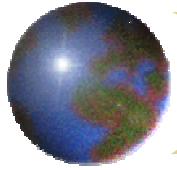
***EC611--Managerial Economics***

**Dr. Savvas C Savvides, European University Cyprus**



# Marketing Research Approaches (1)

- ✦ **Consumer Surveys:** Questioning a sample of consumers to gauge their response to changes in the explanatory vbls. ( $P_x$ ,  $P_y$ ,  $I$ , Features, etc)—*not very reliable*. Yet, sometimes they may be the only option!
- ✦ **Observational Research:** This is a supplementary method to the Consumer surveys using product scanners and people meters at homes (e.g. for TV viewing) to observe changes in consumer behaviors.
- ✦ **Consumer Clinics:** “Laboratory”-style experiments, where a sample of consumers are given a budget and are asked to simulate their spending/buying habits in a “simulated” store or “market”. The objective is again to investigate how they respond to changes in  $P$ , Product features and promotional alternatives). This can also be simulated on computer via **“virtual shopping”**.
- ✦ **Market Experiments:** Similar in concept to the consumer clinics, but the experiments are in the actual marketplace. Different markets with similar demographic and socioeconomic characteristics are chosen. In each market only one controlled variable is changed ( $P$ , packaging, promotional “tricks”). They are expensive (large scale), but useful and effective because they are effective in determining best pricing strategies, and for testing different product features and promotional activities.



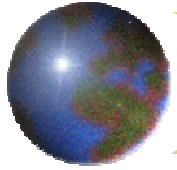
# *Marketing Research Approaches (2)*

## **Virtual Management**

This is a very sophisticated method for simulating consumer behavior using computer models based on the theory of complexity, without relying on actual consumers to simulate actual shopping behavior.

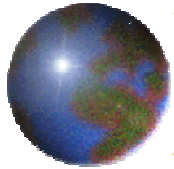
The systems relies on consumer behavior parameters that are programmed in the computer model, based on information from consumer surveys and other database information

This is potentially a very powerful tool that can allow managers to test the impact of various managerial decisions (e.g, changes in P, product features, outlet layout, customer service personnel, etc)



# *Estimation Techniques*

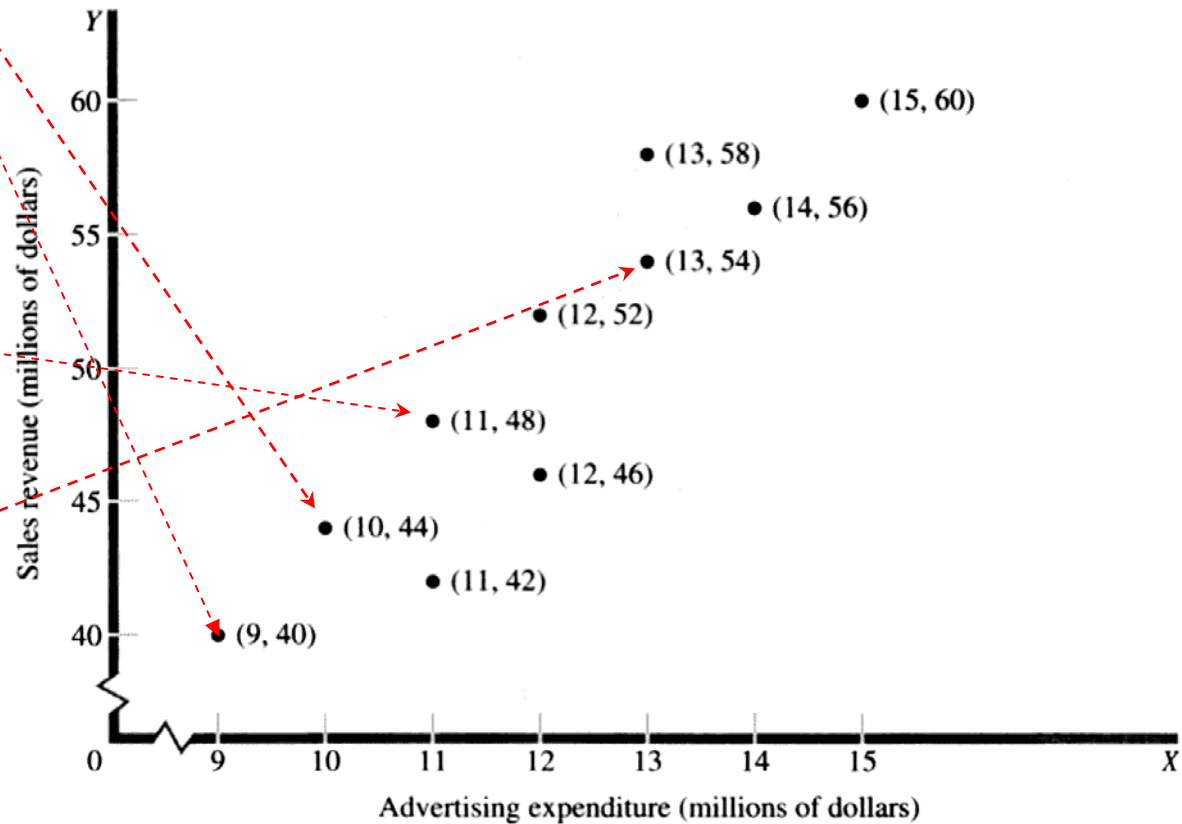
- ✿ These are various quantitative methods to find the exact relationship between the dependent variable and the independent variable(s).
- ✿ The most common method is *regression analysis*
  - ▣ Simple (bivariate) Regression:  $Y = a + bX$
  - ▣ Multiple Regression:  $Y = a + bX_1 + c X_2 + dX_3 + \dots$

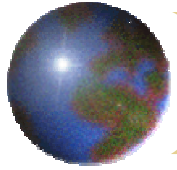


# Regression Analysis

|      | Adv. | Sales |
|------|------|-------|
| Year | X    | Y     |
| 1    | 10   | 44    |
| 2    | 9    | 40    |
| 3    | 11   | 42    |
| 4    | 12   | 46    |
| 5    | 11   | 48    |
| 6    | 12   | 52    |
| 7    | 13   | 54    |
| 8    | 13   | 58    |
| 9    | 14   | 56    |
| 10   | 15   | 60    |

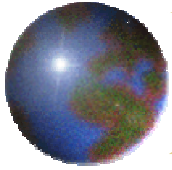
## Scatter Diagram



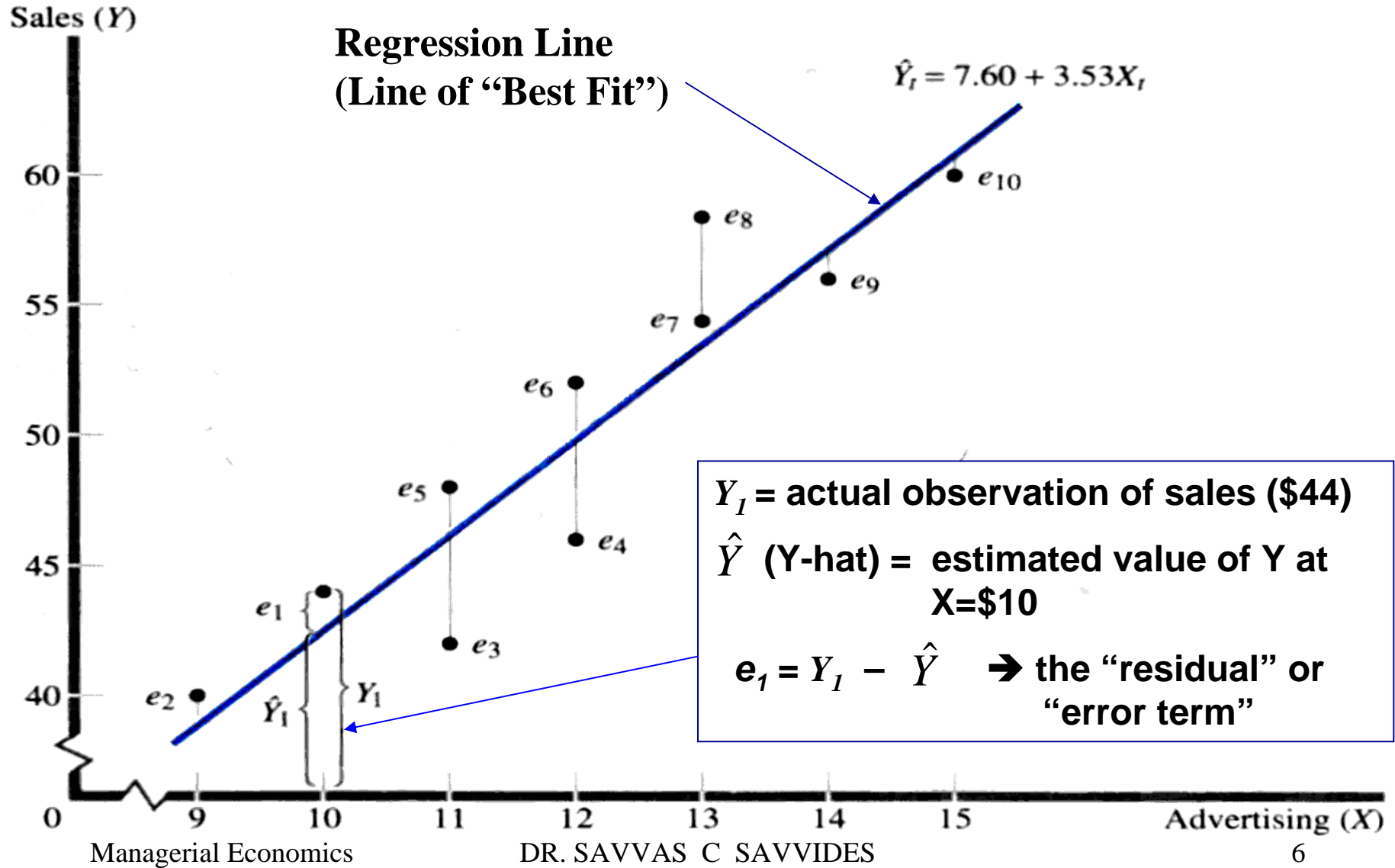


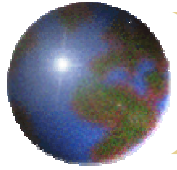
# Regression Analysis

- **Regression Analysis**: This is the statistical technique used to obtain an “imaginary” line that best fits the data points (the **Regression Line** or Line of Best Fit). It is used to:
  - derive estimates of the “slope coefficients” (a, b, c, ...) —estimates of how changes in each independent (explanatory) variable impacts on the dependent variable
  - Conduct test of significance (hypothesis testing)
  - Construct confidence intervals
  - Test for the overall explanatory power of the regression
- **Ordinary Least Squares (OLS)**: This is the method which minimizes the sum of the squared vertical deviations or residuals ( $e_t$ ) of each actual observation point from the regression line.



# Regression Analysis





## Ordinary Least Squares (OLS)

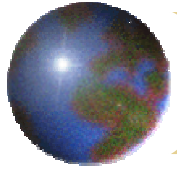
Let's start with a simple two-variable model (say the sales-advertising model examined before) shown in the general form below:

$$Y_t = a + bX_t + e_t$$

Our objective is to find estimates of the slope coefficients ( $a$  and  $b$ ) of the regression line

$Y_t = a + bX_t$  where  $\hat{Y}_t$ ,  $\hat{a}$  and  $\hat{b}$  are the estimates of the dependent variable, the intercept and the slope, respectively.

**The error term (deviation) of the actual Y from its estimated value is given by the equation  $\rightarrow e_t = Y_t - \hat{Y}_t$**



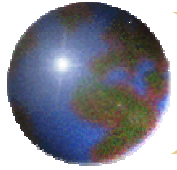
## Ordinary Least Squares (OLS)

**Recall that the OLS is the method which minimizes the sum of the squared vertical deviations or residuals ( $e_t$ ) of each actual observation point from the regression line.**

**It enables us to determine the slope ( $\hat{b}$ ) and intercept ( $\hat{a}$ ) that minimize the sum of these squared errors.**

**The formula for deriving the sum of squared residuals is:**

$$\sum_{t=1}^n e_t^2 = \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 = \sum_{t=1}^n (Y_t - a - bX_t)^2$$



# Ordinary Least Squares (OLS)

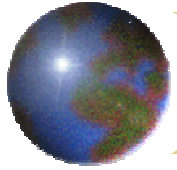
## Estimation Procedure

The estimated value  
of the b coefficient is:

$$b = \frac{\sum_{t=1}^n (X_t - \bar{X})(Y_t - \bar{Y})}{\sum_{t=1}^n (X_t - \bar{X})^2}$$

The estimated value  
of the a coefficient is:

$$a = \bar{Y} - b\bar{X}$$



# Ordinary Least Squares (OLS)

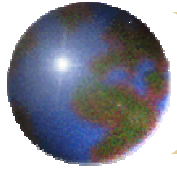
## Estimation Example

| <i>Time</i>       | $X_t$ | $Y_t$ | $X_t - \bar{X}$ | $Y_t - \bar{Y}$ | $(X_t - \bar{X})(Y_t - \bar{Y})$ | $(X_t - \bar{X})^2$ |
|-------------------|-------|-------|-----------------|-----------------|----------------------------------|---------------------|
| 1                 | 10    | 44    | -2              | -6              | 12                               | 4                   |
| 2                 | 9     | 40    | -3              | -10             | 30                               | 9                   |
| 3                 | 11    | 42    | -1              | -8              | 8                                | 1                   |
| 4                 | 12    | 46    | 0               | -4              | 0                                | 0                   |
| 5                 | 11    | 48    | -1              | -2              | 2                                | 1                   |
| 6                 | 12    | 52    | 0               | 2               | 0                                | 0                   |
| 7                 | 13    | 54    | 1               | 4               | 4                                | 1                   |
| 8                 | 13    | 58    | 1               | 8               | 8                                | 1                   |
| 9                 | 14    | 56    | 2               | 6               | 12                               | 4                   |
| 10                | 15    | 60    | 3               | 10              | 30                               | 9                   |
| $\sum_{t=1}^{10}$ | 120   | 500   |                 |                 | 106                              | 30                  |

$n = 10$        $\bar{X} = \frac{\sum_{t=1}^n X_t}{n} = \frac{120}{10} = 12$        $\bar{Y} = \frac{\sum_{t=1}^n Y_t}{n} = \frac{500}{10} = 50$

$a = \bar{Y} - b\bar{X} = 50 - 3.533(12) = 7.60$

$b = \frac{106}{30} = 3.533$



## Ordinary Least Squares (OLS)

Therefore, the equation of the regression line is:

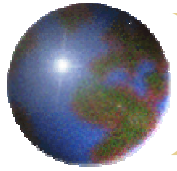
$$Y_t = a + bX_t \quad \rightarrow \quad \hat{Y}_t = 7.60 + 3.53 (X_t)$$

Substituting alternative observed values for  $X_t$  we derive corresponding values for  $\hat{Y}_t$

Plotting these values for  $X_t$  and  $\hat{Y}_t$  we obtain the regression line.

For example, with Advertising ( $X_t$ ) = 0,  $\rightarrow \hat{Y}_t = 7.60$

With  $X_t = 10 \rightarrow \hat{Y}_t = 42.90$  (vs. actual of 44)



## Ordinary Least Squares (OLS)

The estimated regression line can be used to estimate the value of Y that may result from a value of X within or near the range of values used to derive the regression line.

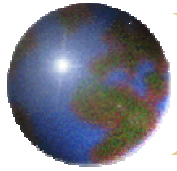
If X's are much higher or lower than this range, then we cannot attach much confidence to the value of the estimated parameters, and ultimately on the estimated value of Y.

The value of the slope coefficient of X ( $\hat{b}$ ) measures the change in the dependent variable (Y) as a result of a change by one unit in the value of X.

It is a marginal measure of the effect of X on Y

$$\rightarrow \hat{b} = \Delta Y / \Delta X \quad \rightarrow \quad \hat{b} = dY/dX \quad (\text{in derivative form})$$

$\rightarrow$  In this sense,  $\hat{b}$  is a *slope!*



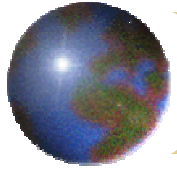
## The Standard Error

The regression results of our example are based on a sample period of ten years. How confident are we that the relationship we found is true for larger or different time periods? To test whether the values of the parameters we estimated indicate that Y and X are “significantly” related (in a statistical sense), we need the Standard Error (or Deviation) of the Slope Estimate (SEE). The formula of the standard error of  $\hat{b}$  is:

$$s_{\hat{b}} = \sqrt{\frac{\sum (Y_t - \hat{Y})^2}{(n - k) \sum (X_t - \bar{X})^2}} = \sqrt{\frac{\sum e_t^2}{(n - k) \sum (X_t - \bar{X})^2}}$$

The smaller the value of this error term in relation to the estimated value of  $\hat{b}$  the larger is our confidence that the value of  $\hat{b}$  is in fact significantly different from zero. In other words, that there is indeed a significant relationship between the actual Y and the actual X.

**The SEE indicates how precisely the model can predict Y.**



# *Tests of Significance*

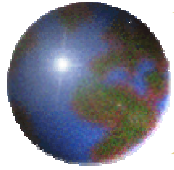
## *Hypothesis Testing*

**Null Hypothesis:** The hypothesis that the value of  $b = 0$ .

**Alternative Hypothesis:** The hypothesis that the value of  $b$  is non-zero.

If we reject the null hypothesis (i.e., that  $b = 0$ ), then we accept the alternative hypothesis that  $b$  is different from zero.

→ This process is called *Hypothesis Testing (or Significance Test)*.



# Tests of Significance

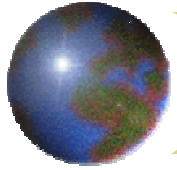
## Example Calculation

| Time | $X_t$ | $Y_t$ | $\hat{Y}_t$ | $e_t = Y_t - \hat{Y}_t$ | $e_t^2 = (Y_t - \hat{Y}_t)^2$ | $(X_t - \bar{X})^2$ |
|------|-------|-------|-------------|-------------------------|-------------------------------|---------------------|
| 1    | 10    | 44    | 42.90       | 1.10                    | 1.2100                        | 4                   |
| 2    | 9     | 40    | 39.37       | 0.63                    | 0.3969                        | 9                   |
| 3    | 11    | 42    | 46.43       | -4.43                   | 19.6249                       | 1                   |
| 4    | 12    | 46    | 49.96       | -3.96                   | 15.6816                       | 0                   |
| 5    | 11    | 48    | 46.43       | 1.57                    | 2.4649                        | 1                   |
| 6    | 12    | 52    | 49.96       | 2.04                    | 4.1616                        | 0                   |
| 7    | 13    | 54    | 53.49       | 0.51                    | 0.2601                        | 1                   |
| 8    | 13    | 58    | 53.49       | 4.51                    | 20.3401                       | 1                   |
| 9    | 14    | 56    | 57.02       | -1.02                   | 1.0404                        | 4                   |
| 10   | 15    | 60    | 60.55       | -0.55                   | 0.3025                        | 9                   |
|      |       |       |             | $\sum_{t=1}^{10}$       | 65.4830                       | 30                  |

$$\sum_{t=1}^n e_t^2 = \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 = 65.4830$$

$$\sum_{t=1}^n (X_t - \bar{X})^2 = 30$$

$$s_b = \sqrt{\frac{\sum (Y_t - \hat{Y}_t)^2}{(n-k) \sum (X_t - \bar{X})^2}} = \sqrt{\frac{65.4830}{(10-2)(30)}} = 0.52$$



# Tests of Significance

## Calculation of the t-Statistic

The basic test of significance is the **t-test**. This is done by dividing the estimated coefficient by its standard error

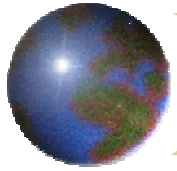
$$t = \frac{b}{s_b} = \frac{3.53}{0.52} = 6.79$$

Degrees of Freedom =  $(n-k) = (10-2) = 8$

Critical Value of t-statistic at 5% level = 2.306

Rule of thumb: t-value of  $>2$

If the estimated coefficient of a variable “passes” the **t-test** we can be confident that X (advertising) truly has an impact on Y (sales). In this case, where  $t=2.306$ , we have a 5% chance that in reality things are not like this.



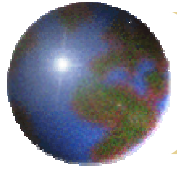
## Goodness of Fit

**Coefficient of Determination ( $R^2$ )**: This is a statistic to test the overall **explanatory power** of the regression, or the “**goodness of fit**” of the observed data around the regression line.

It shows how well the regression explains changes in the value of the Y.

It is defined as the proportion of total variation in Y that is explained by the full set of independent variables. This is shown by the formula:

$$R^2 = \frac{\text{Explained Variation}}{\text{Total Variation}} = \frac{\sum (Y - \bar{Y})^2}{\sum (Y_t - \bar{Y})^2}$$



## Goodness of Fit & Correlation

**$R^2$  can range from zero to 1.** A value of 1 indicates that all the variation in Y is explained by variations in the explanatory variables.

The closer the observed data points fall on the regression line, the smaller the error terms (deviations);  $\rightarrow$  and the greater is the proportion of the variation in Y explained by the variation in X, the larger the value of  $R^2$ . The  $R^2$  for the sales-advertising example is calculated as:

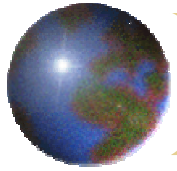
$$R^2 = \frac{373.84}{440.00} = 0.85$$

**Coefficient of Correlation:** This is a measure of the degree of association or correlation (co-variation) between X and Y

$$r = \sqrt{R^2}$$

$$\boxed{-1 \leq r \leq 1}$$

In our example  $r$  is calculated as:  $r = \sqrt{0.85} = 0.92$  This means that X and Y vary (or move) together 92% of the time.



# Regression – An Example

## Estimating the Consumption Function of Cyprus

$Y_t = a + bX_t$  where  $Y_t$  is consumption, and  $X_t$  is income.

$b$  is the slope coefficient (or Marginal Propensity to Consume) which measure the impact (change) on Consumption per unit change in Income

$a$  is the vertical intercept (Consumption at zero Income)—the constant!

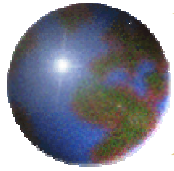
Regressing Consumption on Income by OLS we get the following results:

$$\hat{Y}_t = -14.85 + 0.66 X_t \quad R^2 = 0.9948 \quad n=31$$

t-value (74.45)  $\rightarrow t = \frac{b}{s_b} = 0.66 / 0.00881 = 74.447$

The value of  $\hat{b} = 0.66$  means that (on average) Cypriots spend (consume) 66% of any change in income. A £1m  $\Delta$  in income leads to £0.66m  $\Delta$  in Cons.

Since the t-value of 74.45 is much greater than the **critical value of t (2.045 with n-k = 29)**, we conclude that  $\hat{b}$  is statistically significant (5% level)



# Multiple Regression Analysis

*Model:* 
$$Y = a + b_1X_1 + b_2X_2 + \dots + b_kX_k$$

To estimate multiple regression equations, we use essentially the same procedure as in the simple two-variable case. Here, however, the complexity of the calculations necessitate the use of the computer.

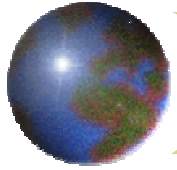
## Analysis of Variance and F Statistic

$$F = \frac{\text{Explained Variation} / (k - 1)}{\text{Unexplained Variation} / (n - k)}$$

$$F = \frac{R^2 / (k - 1)}{(1 - R^2) / (n - k)}$$

The F-stat is used to test the hypothesis that the variation in the independent vbls. as a set (the X's) explains a significant variation in the dependent vbl. Y.

*Null Hypothesis: All regression coefficients are equal to zero.*



# Multiple Regression Analysis

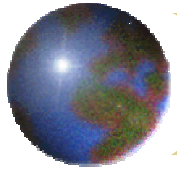
## Adjusted Coefficient of Determination

$R^2$  may be artificially high for small samples.

For meaningful, reliable estimates, the sample size must be sufficiently large (usually over 25-30 observations) to get enough degrees of freedom ( $n-k$ ).

Because  $R^2$  always approaches 100% as  $(n-K) \rightarrow 0$ , statisticians use a method for adjusting  $R^2$  to account for the degrees of freedom. The formula for the adjusted  $R^2$  (denoted as  $\bar{R}^2$ ) is the following:

$$\bar{R}^2 = 1 - (1 - R^2) \frac{(n - 1)}{(n - k)}$$



# Multiple Regression – Example (1)

## The Aggregate Consumption Function of Cyprus Revisited

### (1) Specifying the Consumption Equation

$$C_t = a + b_1 Y + b_2 C_{t-1} + b_3 P + b_4 \text{DUMMY}_{1974}$$

$Y$  is Income (measured by GDP),  $C_{t-1}$  is previous period's private consumption spending (to capture habit-forming behavior),  $P$  stands for inflation (as measured by the Retail Price Index), and  $\text{DUMMY}_{1974}$  is a dummy variable to capture the impact of the Turkish invasion

### (2) Estimating the Consumption Equation

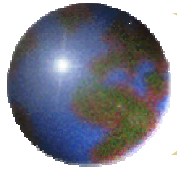
Regressing Consumption on Income, Inflation and Consumption lagged one period by OLS we get:

$$C_t = - 22.55 + 0.25 Y + 0.67 C_{t-1} + 5.59 P - 92.2 \text{DUMMY}_{1974}$$

( 4.13)      (6.78)      (1.57)      (-2.67)      Critical t-value = 2.056

$R^2 = 0.9982$        $R^2\text{-bar} = 0.9979$        $n=31$        $n-1= 4$        $n-k = 26$

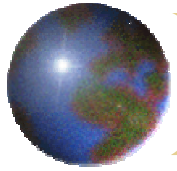
$F\text{-stat} = 3624.5$       Critical F-value (3, 27) = 2.99



## Multiple Regression – Example (2)

### Interpreting the Results

- $R^2 = 0.9982$ . This means that almost all (100%) of the variation in C is explained by  $\Delta$ 's in Y, P &  $C_{t-1}$
- The value of 0.67, the coefficient of  $C_{t-1}$ , means that for every £1m spent last year leads Cypriots to spend £0.67m on current Consumption. Also, 1% increase in inflation leads to an increase in C of £5,59 million. However, since the t-ratio is below 2, we are not statistically confident that this is true.
- The betas for Y and  $C_{t-1}$  are statistically significant (at the 5% level) > **critical value of t (2.05)**
- High value of  $F \rightarrow$  indicates that the regression as a whole explains a statistically significant portion of the variation in C. The **critical value of F (3, 27) = 2.99**



# *Multiple Regression – Example (3)*

## Forecasting Consumption in Cyprus

$$C_t = - 22.55 + 0.25 Y + 0.67C_{t-1} + 5.59P$$

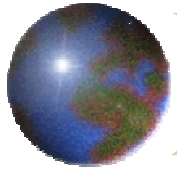
Let's say that the Government estimates that private final consumption expenditures (excluding government consumption expenditures) last year (2003) were £4800 million, this year (2004) Gross Domestic Product is expected to reach £7200 million and this year's inflation rate is expected to be 3%.

When we substitute these values in the estimated equation for Consumption we get:

$$\begin{aligned} C_{2004} &= -22.55 + 0.25 (7200) + 0.67 (4800) + 5.59 (0.03) \\ &= \text{£}5161 \text{ million (for a 7.5\% growth)} \end{aligned}$$

$$\text{Income Elasticity} = 0.25 (7200 / 5161) = 0.35$$

→ aggregate consumption is regarded as a necessity (!!??)



# Problems in Regression Analysis(1)

## Multicollinearity:

Some explanatory variables “move” together over the same period.

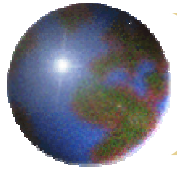
We are not able to distinguish the individual impact on the dependent vbl. All we observe is their combined impact.

- If current income changes in the same way as inflation over time, then we’ll not be able to separate their impact on current consumption
- If advertising is a constant % of total budget, then regressing Sales on Advertising and Total Budget will lead to multicollinearity.

*This leads to low t-values (even though  $R^2$  is high).*

## How do we correct for multicollinearity?

- ⊕ Increase the sample size (collect more data)
- ⊕ Transform the functional form of the equation (e.g deflate data)
- ⊕ Delete one of the collinear variables



# Problems in Regression Analysis(1)

## Autocorrelation:

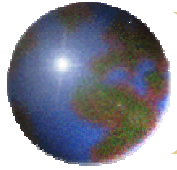
Arises when consecutive error terms have the same sign or change sign frequently. It may be caused by cyclical trends in the X's, from the exclusion of significant vls. or from non-linearities in the data.

Autocorrelation produces artificially high t-values, which erroneously may point to statistically significant parameters. The obtained values for  $R^2$  F-stat will be unreliable in the presence of autocorrelation.

Autocorrelation is detected by the Durbin-Watson statistic, routinely produced by the computer. As a rule of thumb, a value of  $d=2$  is indicative of the absence of autocorrelation.

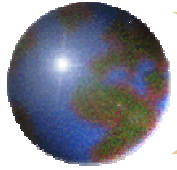
## How do we correct for autocorrelation?

- ⊕ Include time as an additional variable and/or Omit the constant
- ⊕ Transform the functional form of the equation in non-linear form
- ⊕ Include significant variables previously omitted



## *Steps in Demand Estimation*

1. *Specify the Model*: Identify Variables that are believed to be important determinants of the dependent variable (e.g.  $C_t = a + b_1 Y + b_2 C_{t-1} + b_3 P$ ). Use theory and market/empirical knowledge.
2. *Collect Data*: Accurate data are essential to get reliable estimates. Proxy variables: used when data for actual variables are lacking or incomplete or inaccurate (e.g. stock prices may be used as a proxy for wealth). Time-Series or Cross-Sectional data may be used.
3. *Specify Functional Form*: linear, non-linear, log-linear, etc.
4. *Estimate Function*: Run the regression to get estimates for the slope coefficients.
5. *Test the Results*: Run all the necessary tests of significance to test for the statistical significance of each explanatory variable, the explanatory power of the regression, (t-ratio,  $R^2$ , F-test, etc), test for presence of problems and proceed to correct them.



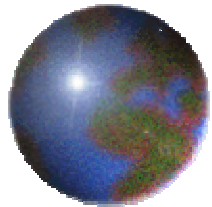
## *Functional Form Specifications*

Linear Function:  $Q_X = a_0 + a_1 P_X + a_2 I + a_3 N + a_4 P_Y + \dots + e$

Power Function:  $Q_X = a(P_X^{b_1})(P_Y^{b_2})$

To estimate a power function using the standard regression techniques, we “linearize” the function; in other words, we transform it into a linear form. To do this, we take the logarithms of the variables as follows:

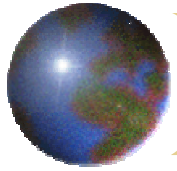
$$\ln Q_X = \ln a + b_1 \ln P_X + b_2 \ln P_Y$$



# **Demand Forecasting**

***EC611--Managerial Economics***

**Dr. Savvas C Savvides, European University Cyprus**



# *Need for Forecasting*

## ✿ Long Range Strategic Planning

- ✿ Corporate Objectives: Profit, market share, ROCE, strategic acquisitions, international expansion, etc

## ✿ Annual Budgeting

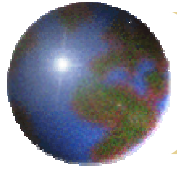
- ✿ Operating Plans: Annual sales, revenues, profits

## ✿ Annual Sales Plans

- ✿ Regional and product specific targets

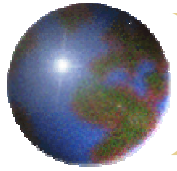
## ✿ Resource Needs Planning

- ✿ HRM, Production, Financing, Marketing, etc



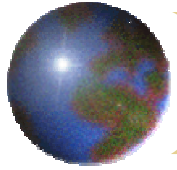
# Selecting Forecasting Techniques

- **Hierarchy of Forecasts:** The level of aggregation desired/required.
  - The macro-level (GNP, disposable income, inflation, plant & equipment, retail sales, etc)
  - Industry sales forecasts
  - Individual firm sales forecasts—by product line. region, etc
- **Criteria used in Selection of Forecasts:** Some are simple and inexpensive, others are complex, time-consuming and expensive. Some are appropriate for ST planning, others for long-range planning / forecasting.
  - ➔ Factors that impact on the selection of technique(s) are:
    - Cost-benefit for developing forecasting model
    - Complexity of behavioural relationships to be forecasted
    - The accuracy of forecasts required
    - The lead time required for making decisions dependent on results of the model



# *Prerequisites for Good Forecasts*

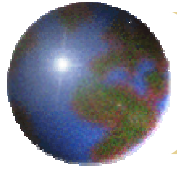
- ✿ **Consistency with other parts of the business**
  - ❑ sales forecasts are as good as the ability of the firm to manufacture and distribute products to the market (on time, good quality, etc)
- ✿ **Based on knowledge of past data**
  - ❑ Exceptions: marketing new products, erratic changes in sales due to exogenous factors
  - ❑ Gut feeling and market experience are relevant here
- ✿ **Consider political and economic factors**
  - ❑ Government policy (taxation, trade, profit repatriation, political instability, FX considerations)
- ✿ **Accuracy vs. Timeliness**
  - ❑ There is usually a trade-off



# *Qualitative Forecasts (1)*

## Survey Techniques -- Macro Level

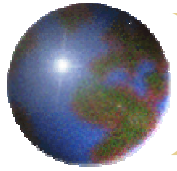
- ❖ **Plant & Equipment Spending Plans:** Surveys (quarterly, bi-annual) for expenditures on fixed assets and R & D (McGraw-Hill, Dept of Commerce, Fortune Magazine, conference Board, etc)
- ❖ **Plans for Inventory Changes & Sales Expectations** (McGraw-Hill, Dept of Commerce, Dunn & Bradstreet, NAPA)
- ❖ **Consumer Expenditure Plans:** Aims at showing trends in consumer income, asset holdings and consumer intentions to purchase specific goods, such as appliances, cars, homes, etc (Census Bureau, U of Michigan).



## *Qualitative Forecasts (2)*

### **Opinion Polls—Micro Level (Sales Forecasting)**

- ❖ **Business Executives:** subjective views of top management (often used in conjunction with quantitative forecasts by trend analysis, etc)
- ❖ **Sales Force:** being close to the market, sales employees have significant insight of the state of consumer psyche, and thus future sales. Care should be used not to confuse “forecasts” with “sales targets”
- ❖ **Consumer Intentions:** mail surveys to estimate consumer intentions of purchasing replacement , upgraded, or complementary (component) products



# *Variation in Time-Series Data*

## ✿ Secular Trend

- ❑ Due to Long-Run changes in economic data series (e.g. population, age distribution, tastes, etc)

## ✿ Cyclical Fluctuations

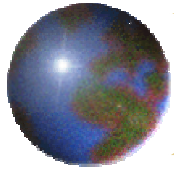
- ❑ Due to major expansions or contractions in economic data (usually >year). Example: housing sales, car sales

## ✿ Seasonal Variation

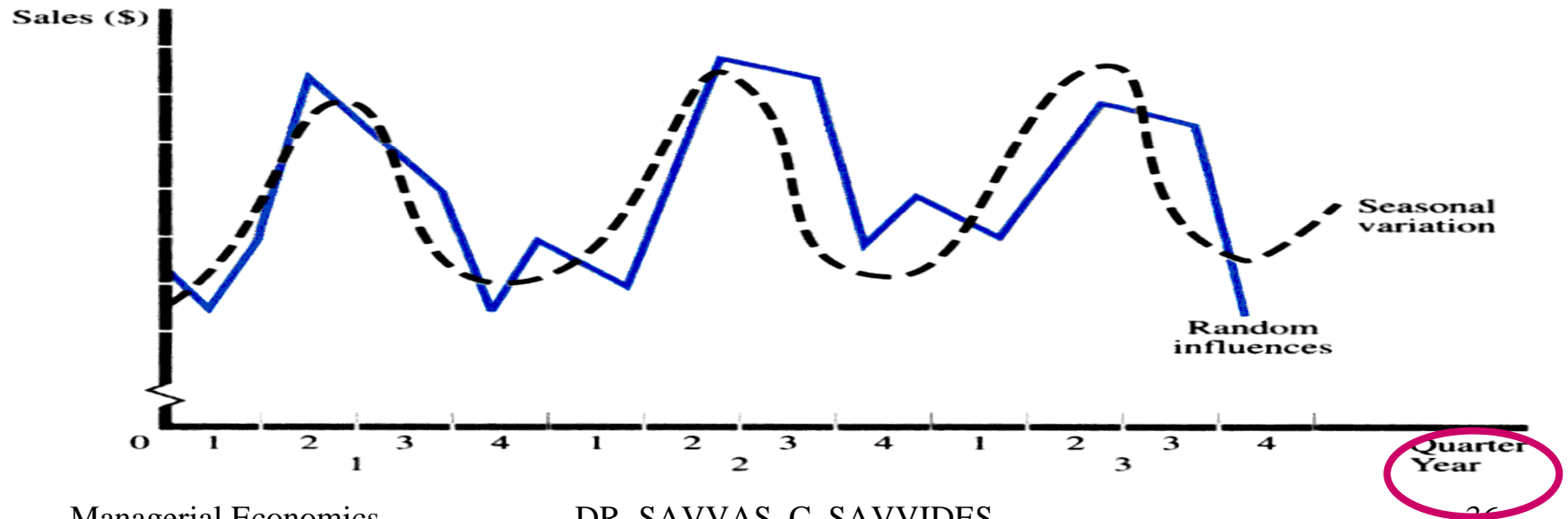
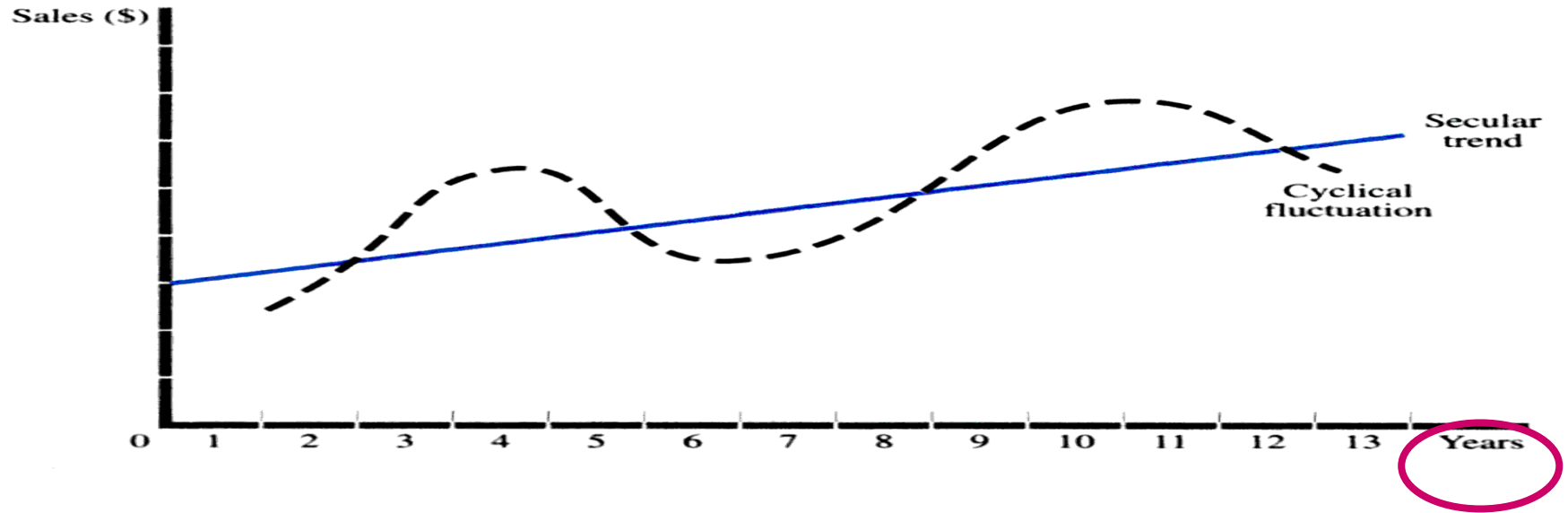
- ❑ Regular/rhythmic fluctuations in sales due to weather, habit, or social custom. Example: Tourism

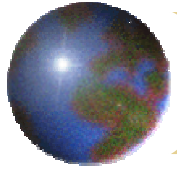
## ✿ Irregular or Random Influences

- ❑ Example: wars, natural disasters, extraordinary government action (imposition of new taxes (VAT), nationalization, trade embargo, etc)



# Time-Series Variation (Graphs)





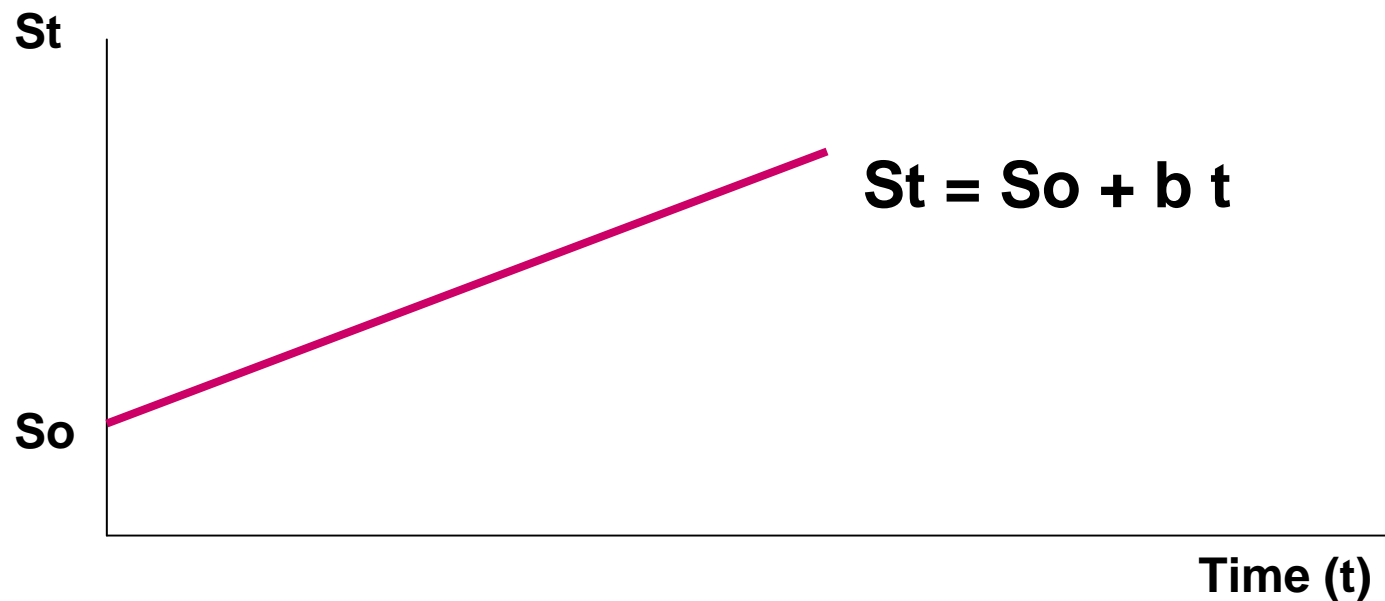
# *Trend Analysis & Projection*

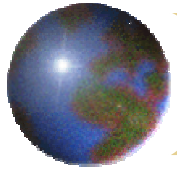
## Linear Trend:

Easy, simple, but inflexible and unsophisticated.

$$S_t = S_0 + b t$$

b = Growth per time period

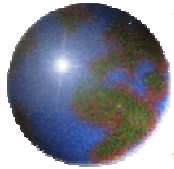




# *Linear Trend Projection—Example (1)*

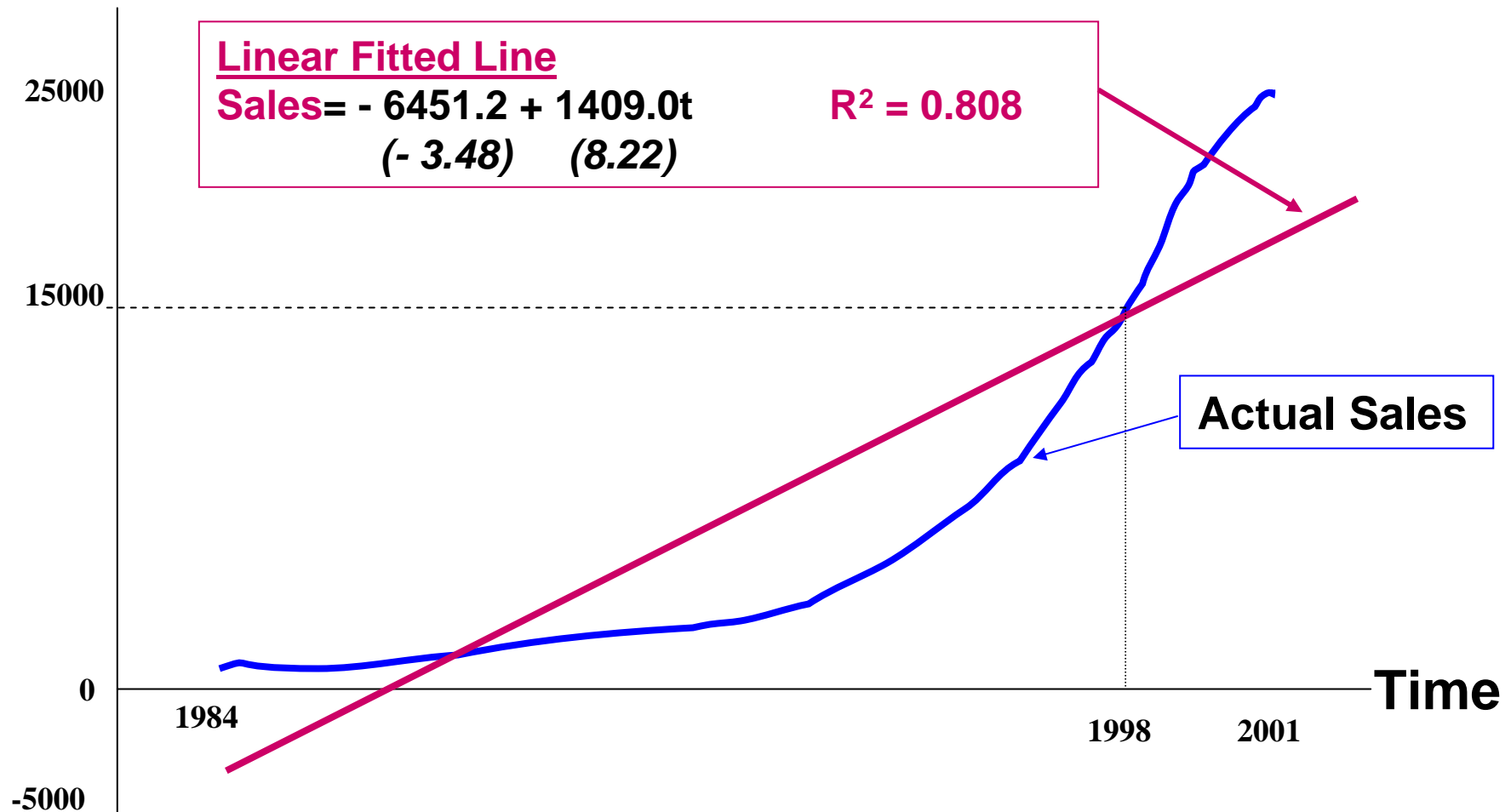
## **Sales Revenue for Microsoft Corp. (1984-2001)--\$ millions**

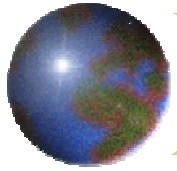
| <b>Year</b> | <b>Time Trend</b> | <b>Actual Sales</b> | <b>Fitted Sales (Linear)</b> |
|-------------|-------------------|---------------------|------------------------------|
| 1984        | 1                 | 100                 | - 5042                       |
| 1985        | 2                 | 140                 | - 3633                       |
| 1986        | 3                 | 202                 | - 2224                       |
| 1987        | 4                 | 346                 | - 815                        |
| 1988        | 5                 | 591                 | 594                          |
| 1989        | 6                 | 804                 | 2003                         |
| 1990        | 7                 | 1183                | 3412                         |
| 1991        | 8                 | 1843                | 4821                         |
| 1992        | 9                 | 2759                | 6223                         |
| 1993        | 10                | 3753                | 7639                         |
| 1994        | 11                | 4649                | 9048                         |
| 1995        | 12                | 5937                | 10457                        |
| 1996        | 13                | 8671                | 11866                        |
| 1997        | 14                | 11358               | 13275                        |
| 1998        | 15                | 14484               | 14684                        |
| 1999        | 16                | 19747               | 16093                        |
| 2000        | 17                | 22956               | 17502                        |
| 2001        | 18                | 25296               | 18911                        |



# Linear Trend Projection—Example (2)

Sales Revenue for Microsoft Corp. (1984-2001)--\$ millions





## Linear Trend Projection—Example (3)

### Forecasts of Sales Revenue for Microsoft Corp. (2002-2005)

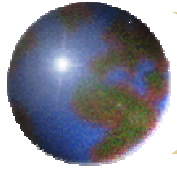
| Year | Time Period | Trend Line               | Fitted Sales (Linear) |
|------|-------------|--------------------------|-----------------------|
| 2002 | 19          | $S = -6451.2 + 1409(19)$ | 20,320                |
| 2003 | 20          | $S = -6451.2 + 1409(20)$ | 21,729                |
| 2004 | 21          | $S = -6451.2 + 1409(21)$ | 23,138                |
| 2005 | 22          | $S = -6451.2 + 1409(22)$ | 24,547                |

The linear trend line model implies that sales increase by a constant value each year. In the Microsoft example, the model projects sales revenue to increase by \$1409 million per year.

The projections should not be far from the current period.

From the previous graph, we see that the true trend line for Microsoft sales is *non-linear*. Thus, the forecasts are poor estimates of actual values.

It may be more appropriate to assume that sales increase at a **constant growth factor (rate)** rather than a constant amount (value)



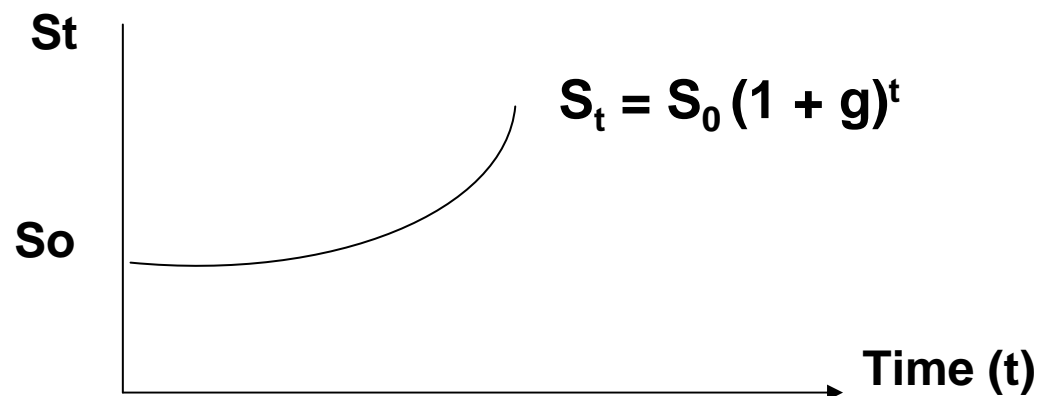
# *Growth Trend Analysis/Projection*

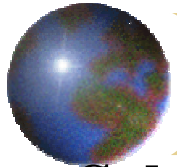
## Constant Growth Rate Trend:

$$S_t = S_0 (1 + g)^t \quad g = \text{Growth rate}$$

To estimate, we “linearise” this behavioral relationship by taking logarithms (natural logs):

$$\ln S_t = \ln S_0 + \ln (1 + g) \times t$$

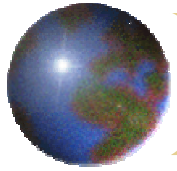




## *Linear Trend Projection—Example (1)*

**Sales Revenue for Microsoft Corp. (1984-2001)--\$ millions**

| <b>Year</b> | <b>Time Trend</b> | <b>Actual Sales</b> | <b>Natural log of Sales</b> |
|-------------|-------------------|---------------------|-----------------------------|
| 1984        | 1                 | 100                 | 4.60517                     |
| 1985        | 2                 | 140                 | 4.941642                    |
| 1986        | 3                 | 202                 | 5.308268                    |
| 1987        | 4                 | 346                 | 5.846439                    |
| 1988        | 5                 | 591                 | 6.381816                    |
| 1989        | 6                 | 804                 | 6.689599                    |
| 1990        | 7                 | 1183                | 7.075809                    |
| 1991        | 8                 | 1843                | 7.51915                     |
| 1992        | 9                 | 2759                | 7.922624                    |
| 1993        | 10                | 3753                | 8.230311                    |
| 1994        | 11                | 4649                | 8.444407                    |
| 1995        | 12                | 5937                | 8.688959                    |
| 1996        | 13                | 8671                | 9.067739                    |
| 1997        | 14                | 11358               | 9.337678                    |
| 1998        | 15                | 14484               | 9.5808                      |
| 1999        | 16                | 19747               | 9.890757                    |
| 2000        | 17                | 22956               | 10.04133                    |
| 2001        | 18                | 25296               | 10.1384                     |



## Growth Trend Projection—Example

Using this “linearized” regression technique to the Microsoft sales data (transformed in logs), we get the following results (t-values in parenthesis):

$$\ln S_t = 4.5695 + 0.33603 t \quad R^2 = 0.983$$

(38.58)      (30.71)

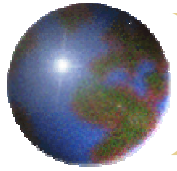
Sales revenue forecasts can be obtained by transforming the equation back into its original form by taking antilogs:

$$S_t = (\text{antilog } 4.57) \times (\text{antilog } 0.336)^t \rightarrow S_t = \$96.38(1.4)^t$$

→ Note that 1.4 in the parenthesis is (1+ g). Therefore, g = .4 or 40%

### Forecasts of Sales Revenue for Microsoft Corp. (2002-2005)

| Year | Time Period (t) | Constant Growth Trend Line | Fitted Sales (Constant Growth) | Fitted Sales (Linear) |
|------|-----------------|----------------------------|--------------------------------|-----------------------|
| 2002 | 19              | $S_t = \$96.38(1.4)^t$     | 57,596                         | 20,320                |
| 2003 | 20              | $S_t = \$96.38(1.4)^t$     | 80,639                         | 21,729                |
| 2004 | 21              | $S_t = \$96.38(1.4)^t$     | 112,896                        | 23,138                |
| 2005 | 22              | $S_t = \$96.38(1.4)^t$     | 158,054                        | 24,547                |

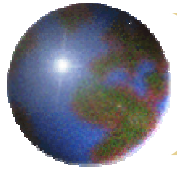


# Seasonal Variation

Often the quantity demanded of (sales) of certain goods does not increase uniformly during the year, but exhibit seasonal patterns. For instance, the demand for ice cream, soft drinks, bottled water, beer, electricity, hotel beds, etc. The sale of all these are higher over the summer months and lower in winter.

## Demand for (Consumption of) Electricity in Cyprus (1998.1 – 2001.4)

| Year / Quarter | Time Period (t) | Quantity (000 kwh) |  | Year / Quarter | Time Period (t) | Quantity (000 kwh) |
|----------------|-----------------|--------------------|--|----------------|-----------------|--------------------|
| 1998.1         | 1               | 2330               |  | 2000.1         | 9               | 2660               |
| 1998.2         | 2               | 2930               |  | 2000.2         | 10              | 3495               |
| 1998.3         | 3               | 2885               |  | 2000.3         | 11              | 3250               |
| 1998.4         | 4               | 2640               |  | 2000.4         | 12              | 2850               |
|                |                 |                    |  |                |                 |                    |
| 1999.1         | 5               | 2470               |  | 2001.1         | 13              | 2775               |
| 1999.2         | 6               | 3245               |  | 2001.2         | 14              | 3640               |
| 1999.3         | 7               | 3050               |  | 2001.3         | 15              | 3350               |
| 1999.4         | 8               | 2800               |  | 2001.4         | 16              | 2950               |



# Seasonal Variation

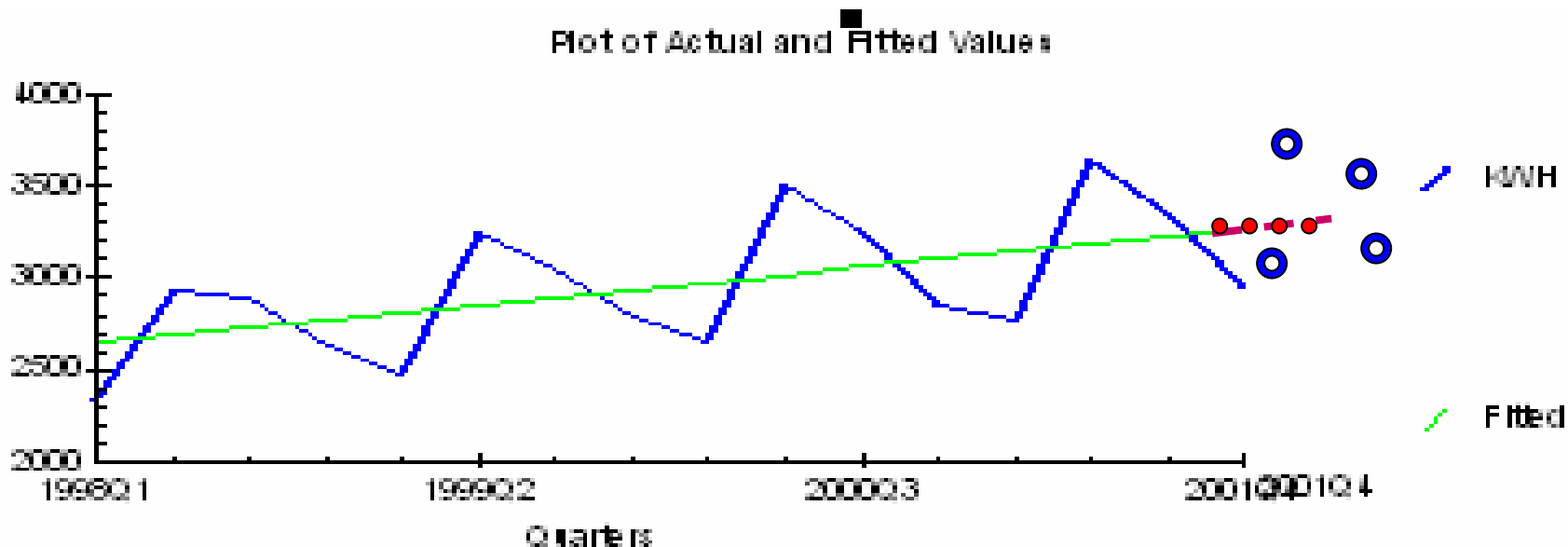
Regressing electricity with linear time trend (t: 1 to 16), we get:

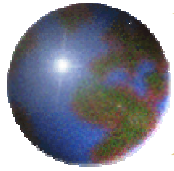
$$S_t = 2605.5 + 41.412t \quad R^2 = 0.294 \quad (\text{t-ratio of } t = 2.42)$$

Forecasting into the following four quarters (2002) we get:

$$S_{17} = \underline{3309.5} \quad S_{18} = \underline{3350.9} \quad S_{19} = \underline{3392.3} \quad S_{20} = \underline{3433.7}$$

We see that the forecasts take into consideration only the long term trend factor in the data. Yet, the data show strong seasonal variation. One way to improve the forecasts is to adjust them with the ratio-to-trend method:





# Seasonal Variation--Example

**Ratio-to-Trend Method**: **Seasonality Ratio** = Actual / Trend Forecast

**Seasonal Adjustment** = Average of Ratios for Each Seasonal Period

**Seasonally Adjusted Forecast** = Trend Forecast \* Seasonal Adjustment

**Calculation for Quarter 17 (2002.1)**

| Year                       | Time Period (t) | Trend Forecast<br>$S_t = 2605.5 + 41.412t$ | Actual Sales | Ratio         |
|----------------------------|-----------------|--|--------------|---------------|
| 1998.1                     | 1               | 2646.9                                     | 2330         | 0.8803        |
| 1999.1                     | 5               | 2812.6                                     | 2470         | 0.8782        |
| 2000.1                     | 9               | 2978.2                                     | 2660         | 0.8932        |
| 2001.1                     | 13              | 3143.9                                     | 2775         | 0.8827        |
| <b>Seasonal Adjustment</b> |                 |  |              | <b>0.8836</b> |

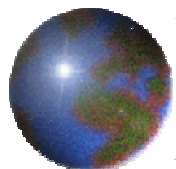
**Seasonally Adjusted Forecast for:**

2002.1 (t =17) = (3309.5)(0.884) = 2924.3 (vs. **3309.5** unadjusted)

2002.2 (t =18) = (3350.9)(1.132) = 3792.5 (vs. **3350.9** unadjusted)

2002.3 (t =19) = (3392.3)(1.053) = 3570.6 (vs. **3392.3** unadjusted)

2002.4 (t =20) = (3433.7)(0.932) = 3199.7 (vs. **3433.7** unadjusted)



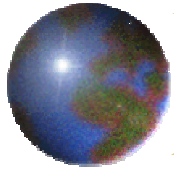
# Seasonal Variation—Dummy Variables

Similar results to the ratio-to-trend method can be obtained by using seasonal dummy variables. Regressing electricity with dummy variables, we get the following results:

```

Ordinary Least Squares Estimation (Seasonal Adjustment with Dummy Variables)
*****
Dependent variable is KWH
16 observations used for estimation from 1998Q1 to 2001Q4
*****
Regressor      Coefficient      Standard Error      T-Ratio[Prob]
C              2405.0           56.4123             42.6326[.000]
D1             -129.7500        54.6614             -2.3737[.037]
D2             598.5000         53.8467             11.1149[.000]
D3             364.2500         53.3519             6.8273[.000]
TIME           40.5000          4.2047              9.6320[.000]
*****
R-Squared      .96860           R-Bar-Squared      .95718
S.E. of Regression  75.2164         F-stat F( 4, 11)    84.8329[.000]
Mean of Dependent Variable  2957.5         S.D. of Dependent Variable  363.5015
Residual Sum of Squares  62232.5         Equation Log-likelihood  -88.8314
Akaike Info. Criterion  -93.8314        Schwarz Bayesian Criterion  -95.7628
DW-statistic   1.7476

```



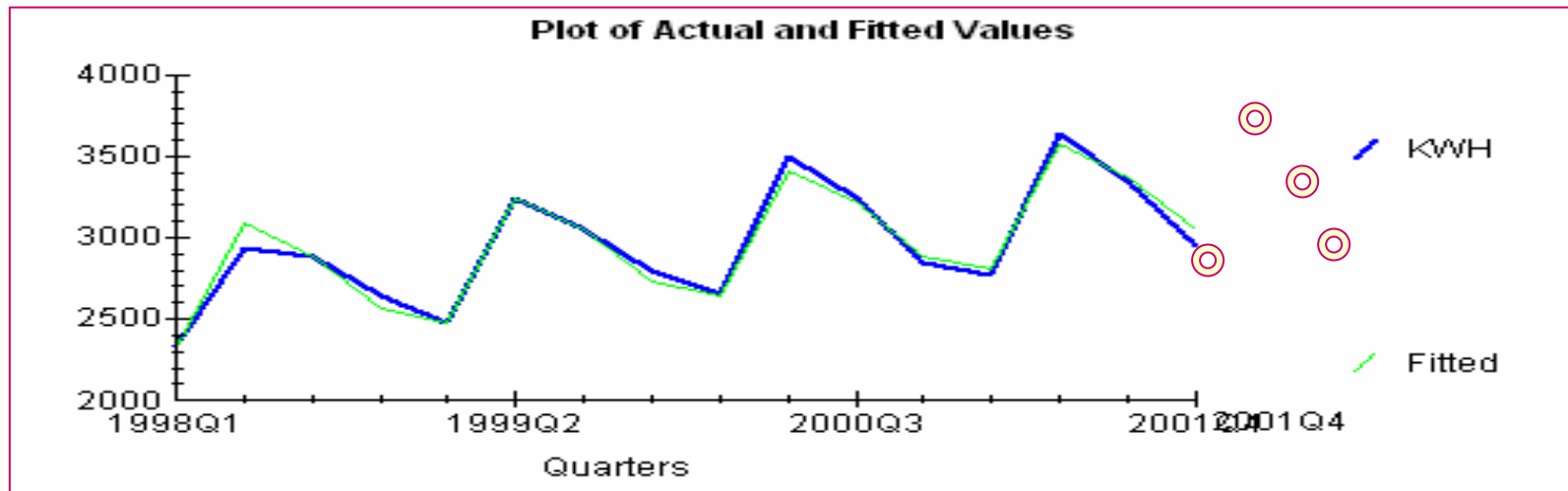
# Seasonal Variation—Dummy Variables

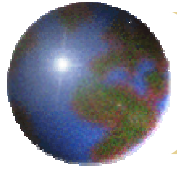
$$\text{KWH} = 2405 - 129.75D1 + 598.5 D2 + 364.25D3 + 40.5 t \quad R^2 = 0.969$$

(-2.38)      (11.11)      (6.83)      (9.63)

We can now use these regression results to forecast electricity consumption for each of the four quarters in 2002, as follows:

| <u>Adjusted forecasts using dummy variables:</u>           | <u>Ratio-to-Trend</u> | <u>Unadjusted</u> |
|--|-----------------------|-------------------|
| 2002.1 (t =17) = 2405 – 129.75 +40.5 (17) = <u>2963.75</u> | 2924.3                | 3309.5            |
| 2002.2 (t =18) = 2405 + 598.50 +40.5 (18) = <u>3732.5</u>  | 3792.5                | 3350.9            |
| 2002.3 (t =19) = 2405 + 364.25 +40.5 (19) = <u>3539.75</u> | 3570.6                | 3392.3            |
| 2002.4 (t =20) = 2405 +40.5 (20) = <u>3210.0</u>           | 3199.7                | 3433.7            |





# Smoothing Techniques (1)

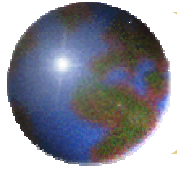
## Moving Average Forecast

This is another naïve forecasting technique. Useful when there are no regularities in the data series (random variation). The moving average forecast is the average of data from  $w$  periods prior to the forecast data point.

For example, the twelve-month moving average forecast for sales of a product for March 2004 is the average of sales for the previous twelve months (March 2003–February 2004).

The greater the number of periods in the moving average, the greater is the smoothing effect since each new observation has a small weight. This is useful for erratic/random time-series data.

$$F_t = \sum_{i=1}^w \frac{A_{t-i}}{w}$$



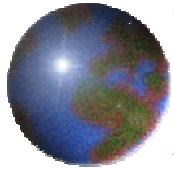
## Smoothing Techniques (2)

**Exponential smoothing Technique:** The forecast for period  $t+1$  (that is  $F_{t+1}$ ) is the **weighted average** of the actual value from the prior period **and** the forecast ( $F_t$ ).

$$F_{t+1} = wA_t + (1 - w)F_t \quad 0 \leq w \leq 1$$

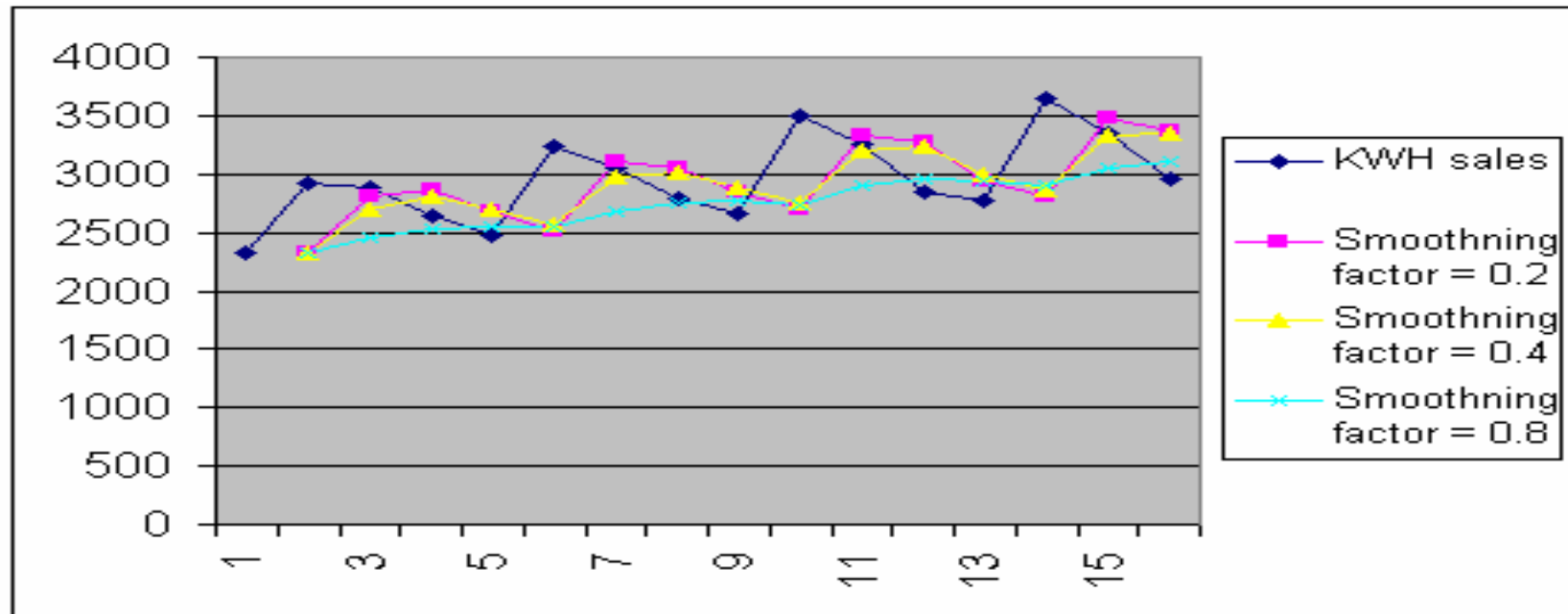
**Root Mean Square Error :** It measures the Accuracy of a Forecasting Method. The lower the RMSE, the better the forecast.

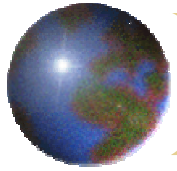
$$RMSE = \sqrt{\frac{\sum (A_t - F_t)^2}{n}}$$



# Exponential Smoothing

| Quarter | KWH sales | Smoothing factor = 0.2 | Smoothing factor = 0.4 | Smoothing factor = 0.8 |
|---------|-----------|------------------------|------------------------|------------------------|
| 1       | 2330      | #N/A                   | #N/A                   | #N/A                   |
| 2       | 2930      | 2330                   | 2330                   | 2330                   |
| 3       | 2880      | 2810                   | 2690                   | 2450                   |
| 4       | 2640      | 2870                   | 2807                   | 2537                   |
| 5       | 2470      | 2686                   | 2707                   | 2558                   |
| 6       | 3240      | 2513                   | 2565                   | 2540                   |
| 7       | 3060      | 3099                   | 2973                   | 2681                   |
| 8       | 2800      | 3060                   | 3019                   | 2755                   |
| 9       | 2660      | 2852                   | 2888                   | 2764                   |
| 10      | 3490      | 2698                   | 2751                   | 2743                   |
| 11      | 3260      | 3336                   | 3197                   | 2893                   |
| 12      | 2860      | 3267                   | 3229                   | 2965                   |
| 13      | 2770      | 2933                   | 3002                   | 2942                   |
| 14      | 3640      | 2807                   | 2866                   | 2908                   |
| 15      | 3360      | 3473                   | 3330                   | 3055                   |
| 16      | 2960      | 3375                   | 3342                   | 3114                   |

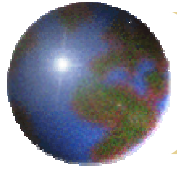




# *Barometric Forecasting Methods*

**Macro-Economic Indicators**: they show the state of the economy and the direction of short-term changes or turning points in business cycles. Such series of economic indicators are produced regularly by the NBER and the Conference Board.

- ✚ **Leading Indicators**: show where general business activity is heading (examples: housing starts, stock market index, money supply, index of consumer expectations, etc)
- ✚ **Lagging Indicators**: follow (lag) economic activity (e.g.,  $\Delta$ 's in labor costs, interest rate, RPI for services, average duration of unemployment, etc)
- ✚ **Coincident Indicators**: move in line (coincide with) economic activity (e.g., disposable income, industrial production)
- ✚ **Composite Index** (e.g of 10 leading indicators)
- ✚ **Diffusion Index** ("breadth" ( %) of them moving up or down)



# *Econometric Models*

## Single Equation Model of the Demand For Good X

$$Q_X = a_0 + a_1P_X + a_2Y + a_3N + a_4P_S + a_5P_C + a_6A + e$$

$Q_X$  = Quantity of X

$P_S$  = Price of Substitute

$P_X$  = Price of Good X

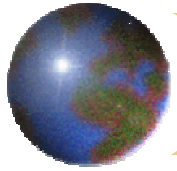
$P_C$  = Price of Complement

Y = Consumer Income

A = Advertising Budget

N = Size of Population

e = Random Error



## *Econometric Methods—Example*

Suppose that we estimate the sales of economy class tickets (in thousands) between New York and London and get the following results (variables are transformed in natural logs):

$$\ln S_t = 2.737 - 1.247 \ln P_t + 1.905 Y_t$$

(- 3.73)                      (10.8)                      R<sup>2</sup> = 0.975

Assume further that an airline company gets estimates for next year that Y (USA income) = \$1480 (billion) and P (prices) = \$550.

→  $\ln(550) = 6.31$  and  $\ln(1480) = 7.3$

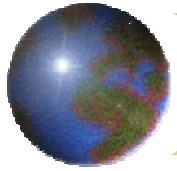
Substituting in the regressed equation above yields:

$$\ln S_{t+1} = 2.737 - 1.247 (6.31) + 1.905 (7.3)$$
$$\ln S_{t+1} = 8.7755$$

Antilog of 8.7755 is 6470. Therefore, sales of airline tickets next year will be 6.47 million.

Since the variables are in logs, the estimated slope coefficients are elasticities.

→ Price elasticity = 1.247    and Income elasticity = 1.905



# *Econometric Models*

## Multiple Equation Model of GNP

$$C_t = a_1 + b_1 GNP_t + u_{1t}$$

$$I_t = a_2 + b_2 \pi_{t-1} + u_{2t}$$

$$GNP_t \equiv C_t + I_t + G_t$$

## Reduced Form Equation

$$GNP_t = \frac{a_1 + a_2}{1 - b_1} + \frac{b_2 \pi_{t-1}}{1 - b_1} + \frac{G_t}{1 - b_1}$$