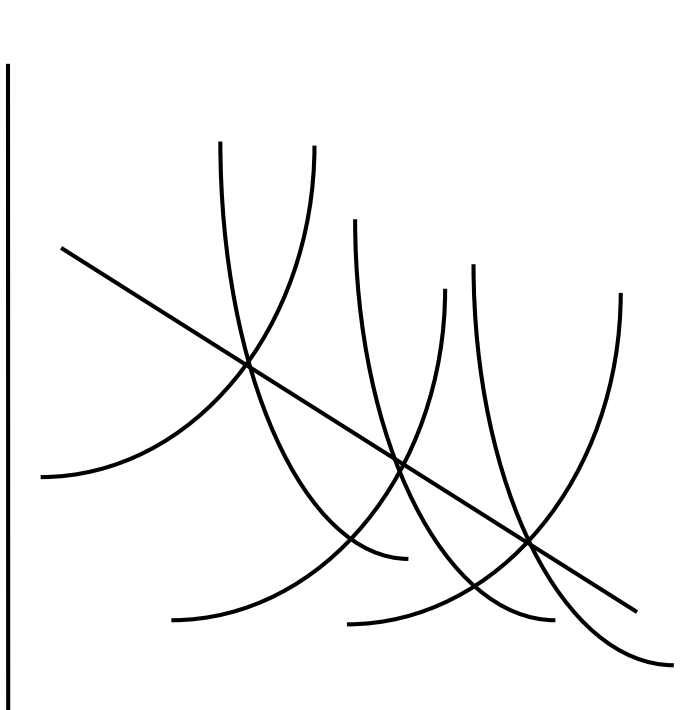
Demand Estimation

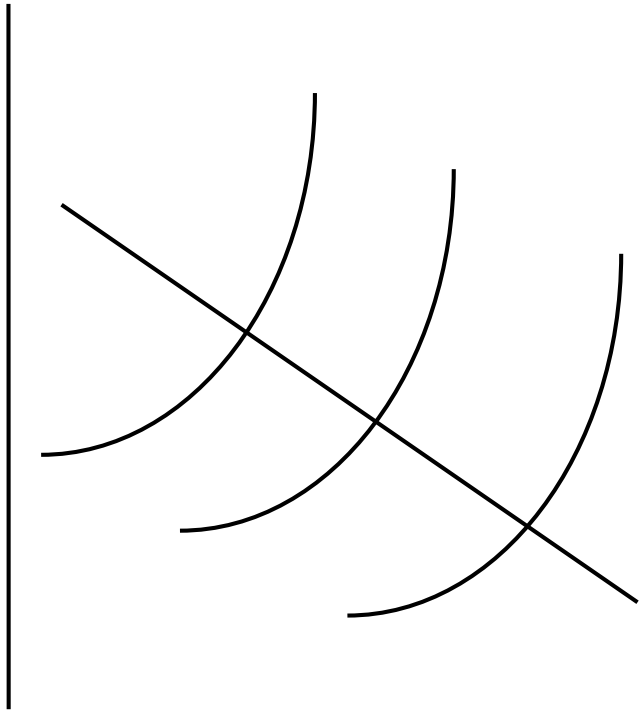
Identification Problem- For our purposes this is the problem incurred by including independent variables that do not effect the dependent variable or exclude independent variables that are relevant.

Difficulties in estimating Demand

It is difficult to estimate all of the variables that may affect demand.

In addition, the interaction of demand and supply can distort any estimate of demand.





Simultaneous relation

Concurrent association. In our example the supply and demand determines the quantity and price of the product. We therefore must estimate supply and demand simultaneously.

This of course make the identification problem even harder since we have to correctly predict which factor affects supply and demand. In addition, we have to deal with factors that may affect both supply and demand.

Interview and Experimental Methods

Consumer Interviews - Questioning customers to estimate demand relations.

Many times customers are unable or unwilling to provide correct answers to surveys.

This is why a market experiment may be a better option.

It can involve the manipulation of prices or different types of advertisement in certain test markets. This way they can see how consumers actually respond to changes in these variables. One problem is that the tester can not control environmental factor that can exogenously effect the demand for the product.

Another method is a laboratory experiment. Participants are given monetary units to buy certain products. Researcher can then manipulate the environment to test how certain change will affect demand.

One problem with type of study is that participants are likely to act differently in the experiment, than they might in the real world. Another problem is the high cost of such experiments.

Using both these method we can estimate point or arc price elasticity. We can also measure cross elasticity. We do however have to be aware of the limitations to estimations of demand.

REGRESSION ANALYSIS

A statistical technique that describes the relation among dependent and independent variables. In other words it describes how one variable is related to another. In multiple variable regression analysis, it will describe a relationship between two variable holding other relative variable constant.

Deterministic Relation
An association that is known with certainty.

Statistical Relation
An imprecise link between two variables.

We must also be careful to differentiate between correlation and causation.

Correlation is the statistical relation between two variables, it however does not tell us if one variable causes movement in the other variable.

Because it is very difficult to differentiate between correlation and causation we must be careful in choosing the variables in regressions. The variables should have a theoretical or logical relationship, if not it is more likely to get variables that are correlated but do not have a causal relationship.

Ex. Population of storks in Norway is positively related with the human birth rate.

Therefore a researcher may conclude that storks deliver babies.

Reverse Causation- the direction of causation may be reversed from what the researcher believes.

Ex. At a party some kids are playing on their parents exercise machine. One of the kids tell the other not to play on the exercise machine because if they use it they would get fat (like their mother).

The child concluded that it is exercise that made one fat.

In both examples the conclusion is wrong.

In the first example the although their is a correlation between the number or storks and the number of newborn humans, their is no causation.

In the second example the kids concluded that the correlation between exercise and body fat, meant that exercise contributed to body fat. The true causation is reversed, it is because her parents were overweight is why the use they exercise machine.

Types of Data

There are 3 types of data broken up by the length of time used. These include time series data and cross sectional data. The third type of data is a combination of the first two and is called panel data.

Time Series Data
A periodic sequence of economic data, usually tracks one data point over time.

Cross Section data
A sample of data taken at a given point in time, has many data points

Panel data
Has many data points within as well as across time periods.

The simplest way to analyze a sample is with a scatter diagram.
Scatter Diagram is a plot of data where the dependent variable is plotted against the independent variable.

With time series data you may want to use time the as independent variable.

With a scatter diagram one can

Y
X

Y
t

Specifying the regression model.

Pick the variables to be included. As stated before they should not be picked randomly, and should be based on logic or theory.

Many times data on the variables you choose will not be available. In these cases we may want to choose a proxy for the variable we have chosen. Other options are to collect the data or estimate the variable..

Functional Form

Linear models - A straight line relation

$Q = b_0 + b_1X_1 + b_2X_2$

b_1 is the amount Q changes when X_1 changes.

Non-linear Models

Multiplicative models A non-linear relationship that involves X variables interactions.

Ex. Cobb Douglas production function.

Least Squares method.

This is a method that fits a line that has the smallest total squared distance from each data point.

In the graph above the line has the smallest total (squared) distance from each data point.

Simple Regression- The relation between one dependent variable and one independent variable.

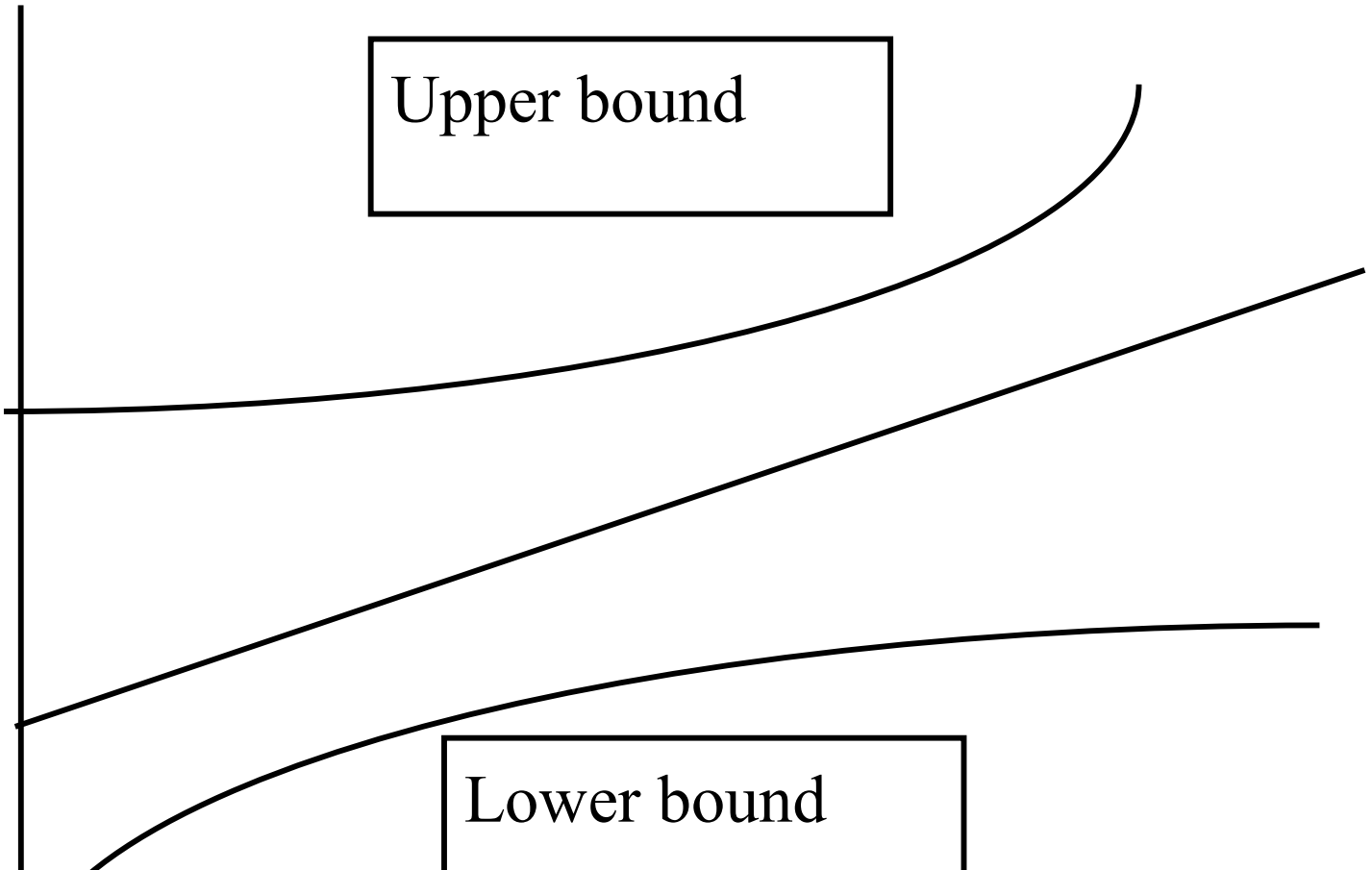
Multiple regression- The relation between an dependent variable and two or more independent variables.

Measures of Regression Model Significance.

Standard Error or the Estimate. (S.E.E.) The standard deviation of the dependent variable after controlling for the influence of all X variables. In other words this is a measure of how well the regression line fits the data. If all the data point fall on the line the S.E.E. will be zero. The farther away they are the bigger the S.E.E will become..

Confidence Intervals

$\ddot{Y} = t^* \text{ S.E.E.}$



We can be confident with some level of probablity that the “true” regression line will fall between the upper bound and the lower bound. The level of probability will determine the level of “t.”

Goodness of fit,

Correlation coefficient (r) - A goodness
for a simple (one independent variable)
regression model. (-1<r<1).

Coefficient of determination (R^2)- A mea
of the goodness of fit for a multiple regre
model; the square of the coefficient of
multiple correlations. (0>R^2>1).

In other word how much of the variation in the dependent variable is explained by the model. R^2 = (Variation Explained by Regression / Total Variation of Y)

Corrected Coefficient of determination (adjusted R^2) - A downward adjustment to R^2 in light of the number of data points and estimated.

adjusted $R^2 = R^2 - [(k-1)/(n-k)](1-R^2)$

Where (n-k) denotes the degrees of freedom (dof) - the # of observation beyond the minimum

F-statistic - A measure of statistical significance for the share of dependent variable variation explained by the regression model.

$$F_{k-1,n-k} = \frac{\frac{(\text{Explained variation})}{(k-1)}}{\frac{\text{Unexplained Variation}}{n-k}}$$

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

t-statistic

This statistic describes the difference between an estimated coefficient and some hypothesized value in terms of "standardized units," or by the number of standard deviations of the coefficient.

$$t_{n-k} = \frac{\hat{b} - b_{null}}{\text{Standard Error of } \hat{b}}$$

two tailed t test - Tests of the b=0 hypothesis.

one-tail t test - tests of direction or comparative magnitude.

Stepwise Regression - An experimental method of independent variable selection based upon XY correlation.

After picking the independent variables that should effect dependent variable, a researcher can add and subtract variables to get the model that has the best fit. In addition, one can change functional form. This should only be done with variable that theoretical or logically affect the dependent variable.

Problems with Regression Models

Multicollinearity - The situation in which two or more independent variables are very highly correlated.

This makes it impossible to identify the degree to which an individual independent variable affects the dependent.

Residual -Error or unexplained variation. Distance of data point from regression line.

Hetroskedaticity - Non-constant variance in the distribution of the residual. .

Serial or Auto Correlation - The residual in of one data point is correlated with other regressions. This is usually only a problem with time series data, and therefore the data points that are correlated are across time.

Durbin Watson Statistic - A measure of the extent of serial correlation. The value the should be as close to 2 as possible. Otherwise the regression has serial correlation.

The Estimate of a firms Demand function is shown below

$$Q_X = -100 - 2.1P_X + 1.6P_Y + 3.2I + 1.5A$$
$$\quad (25.6) \quad (0.81) \quad (1.21) \quad (1.11) \quad (0.35)$$

$R^2 = 0.85$

F-Statistic = 5.13

Standard Errors of the Coefficients are in Parenthesis

Where

Q_X=The Quantity Demanded for X
P_X=The Price of X
P_Y=The Price of a Competing product Y
I=Per Capita income
A=Advertising Expenditures

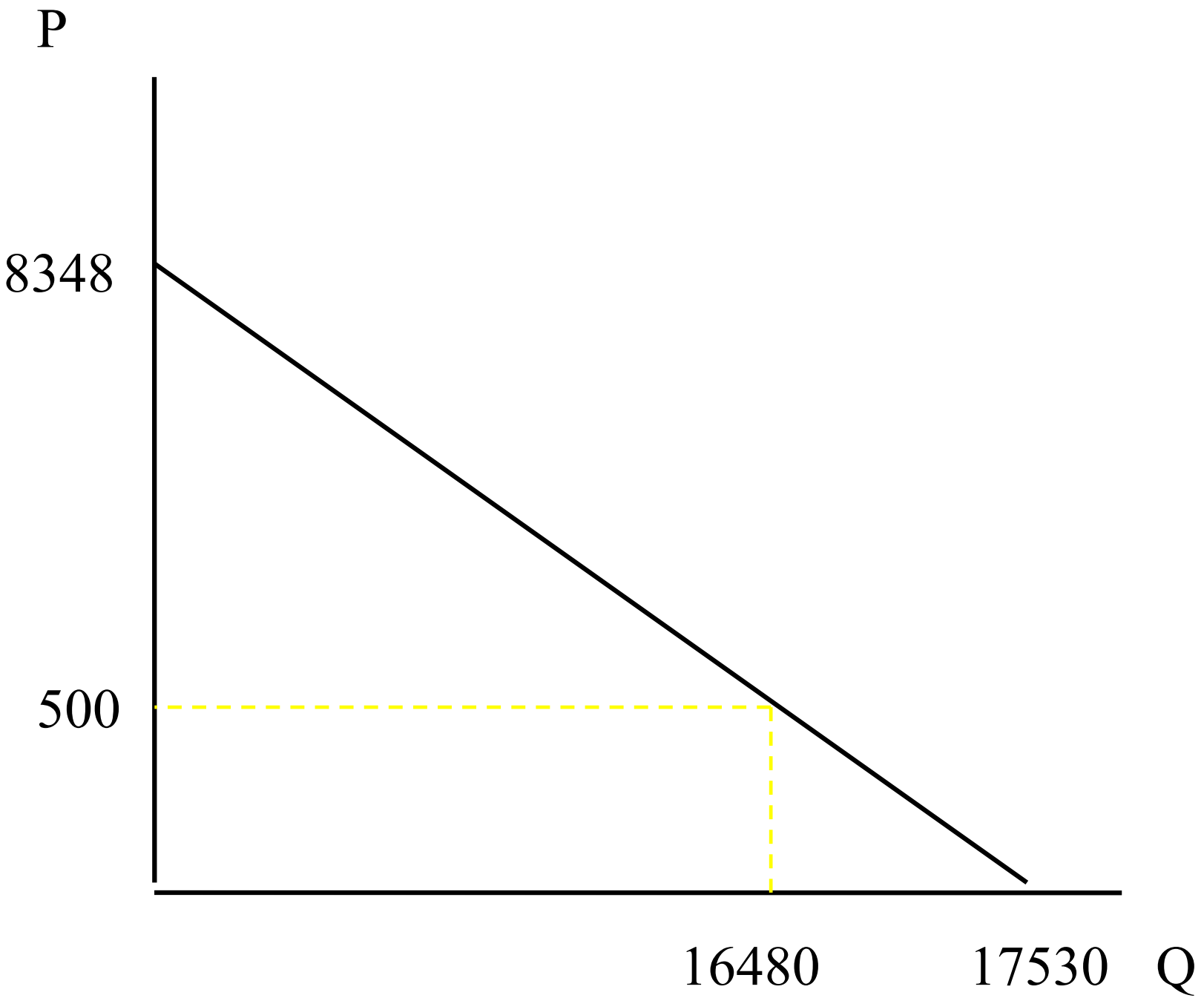
The p-value is 0.005 for the coefficient of A. This means the probability that the t statistic for the regression coefficient of A would be as large (in absolute terms) as it is in this case if in fact A has no effect on Q_x. Interpret this result.

Assume I=5000, P_Y=1000 (in dollars) and A=20 (in thousands of dollars).

If P=500 estimate the quantity demanded for this product.

Assume P is no longer fixed at 500 and now can vary.

Draw the estimated demand curve for this product suing at least 2 points.


P
8348
500
16480
17530
Q

Estimate the t statistic for the coefficient of P_X and I using the alternative (null) hypothesis that the coefficient is not statistically different from zero.

How well does this regression fit the data?

Example: SOFTWARE PACKAGES AND COMPUTER PRINTOUTS

FIGURE 4.7

Minitab Printout of Results of Multiple Regression

```
MTB > regress c1 on 2 predictors in c2 and c3

The regression equation is
C1 = 2.53 + 1.76 C2 - 0.352 C3

Predictor              Coef            Stdev          t-ratio            p
Constant             2.5294           0.2884             8.77        0.000
C2                  1.75805          0.06937            25.34        0.000
C3                 -0.35187          0.07064            -4.98        0.002

s = 0.3702             R-sq = 99.4%     R-sq(adj) = 99.2%

Analysis of Variance

SOURCE          DF              SS              MS              F            p
Regression       2         144.067          72.033          525.72        0.000
Error            6           0.822           0.137
Total            8         144.889

SOURCE          DF          SEQ SS
C2               1         140.667
C3               1           3.399
```

Managerial Economics, 8e
Copyright @ W.W. & Company 2013

INTERPRETING THE OUTPUT OF STATISTICAL SOFTWARE (CONT'D)

FIGURE 4.8

SAS Printout of Results of Multiple Regression

Dependent Variable: C1

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	144.06678	72.03339	525.718	0.0001
Error	6	0.82211	0.13702		
C Total	8	144.88889			

Root MSE	0.37016	R-square	0.9943
Dep Mean	10.88889	Adj R-sq	0.9924
C.V.	3.39944		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for HO: Parameter=0	Prob > |T|
INTERCEP	1	2.529431	0.28842968	8.770	0.0001
C2	1	1.758049	0.06937127	25.343	0.0001
C3	1	−0.351870	0.07064425	−4.981	0.0025

Copyright @ W.W. & Company 2013