Demand Forecasting II
Causal Analysis

Chris Caplice
ESD.260/15.770/1.260 Logistics Systems
Sept 2006

Massachusetts
Institute of
Technology

Agenda

- Forecasting Evaluation
- Use of Causal Models in Forecasting
- Approach and Methods
  - Ordinary Least Squares (OLS) Regression
  - Other Approaches
- Closing Comments on Forecasting

Forecast Evaluation

- How do we determine what is a good forecast?
  - Accuracy - Closeness to actual observations
  - Bias - Persistent tendency to over or under predict
  - Fit versus Forecast – Tradeoff between accuracy to past forecast to usefulness of predictability
  - Forecast Optimality – Error is equal to the random noise distribution
- Combination of art and science
  - Statistically – find a valid model
  - Art – find a model that makes sense

Accuracy and Bias Measures

1. Forecast Error: $e_t = x_t - \hat{x}_t$

2. Mean Deviation:

$$MD = \frac{\sum_{t=1}^{n} e_t}{n}$$

3. Mean Absolute Deviation

$$MAD = \frac{\sum_{t=1}^{n} |e_t|}{n}$$

4. Mean Squared Error:

$$MSE = \frac{\sum_{t=1}^{n} e_t^2}{n}$$

5. Root Mean Squared Error:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n} e_t^2}{n}}$$

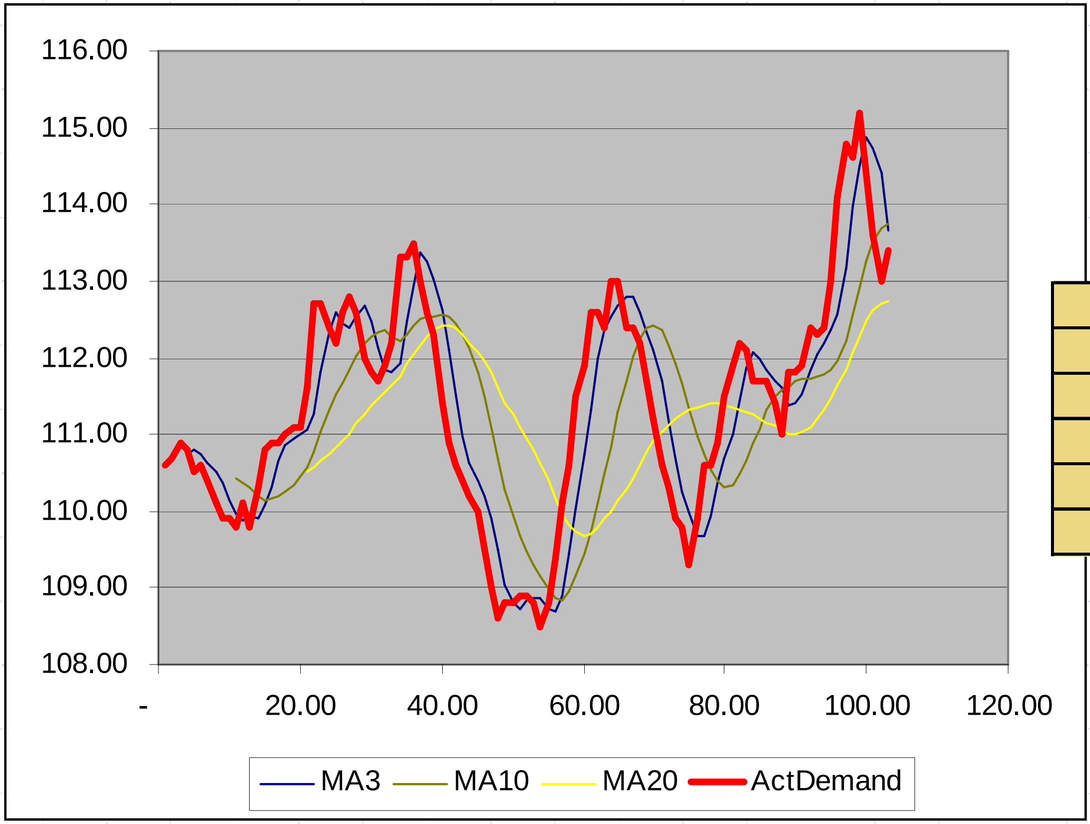
6. Mean Percent Error:

$$MPE = \frac{\sum_{t=1}^{n} \frac{e_t}{D_t}}{n}$$

7. Mean Absolute Percent Error:

$$MAPE = \frac{\sum_{t=1}^{n} \frac{|e_t|}{D_t}}{n}$$

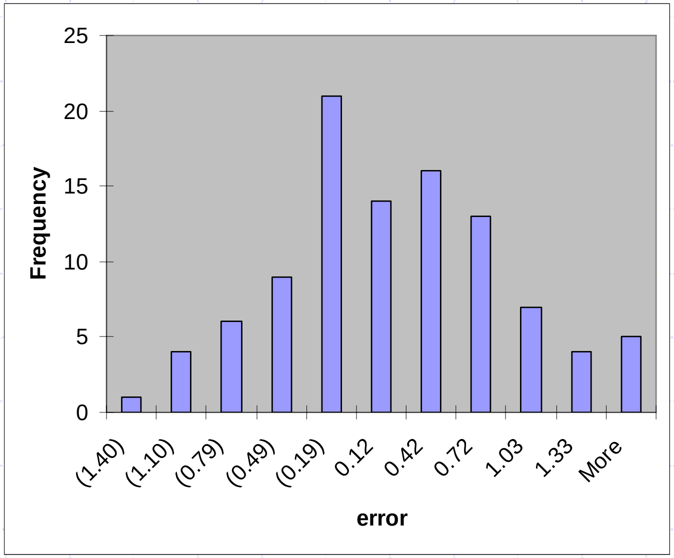
Moving Average Forecasts

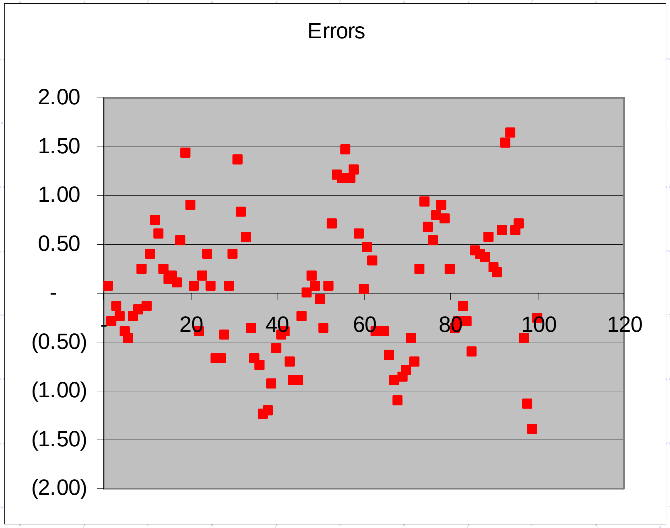


	MA3	MA10	MA20
MD	0.05	0.21	0.35
MAD	0.56	1.07	1.41
MSE	0.47	1.67	2.71
RMSE	0.68	1.29	1.65
MAPE	0.50%	0.96%	1.27%

Analysis of the Forecast

- Are the forecast errors ~N(0,Var(e))?
  - For Moving Averages:
    - What is the expected value of the errors?
    - What is the variance of the errors?
  - From actual observations,
    - Are the observed errors ~N(0,Var(e))?
    - For the MA3 data
      - $\mu_e = 0.05$
      - $\sigma_e = 0.69$
      - $\sigma_D = 1.478$
    - Testing for Normalcy – Chi-Square, Kolmogorov-Smirnov, or other tests





Corrective Actions to Forecasts

- **Measures of Bias**
  - Cumulative Sum of Errors (C_t)
    - Normalize by dividing by RMSE (U_t)
    - U_t should ~0 if unbiased
  - Smoothed Error Tracking Signal (T_t)
    - $T_t = z_t / MAD_t$
    - Where $z_t = \omega e_t + (1-\omega) z_{t-1}$ (smoothing constant)
  - Autocorrelation of forecast Errors
    - Correlation between successive observations
- **Corrective Actions**
  - Adaptive Forecasting
    - Methods where the smoothing coefficients change over time
    - Found (generally) to be no better than standard methods
  - Human Intervention
    - Overrule the model's output – look for reason
    - Rules of thumb: $|T_t| > f$ or $|C_t| > k(RMSE)$ ($f \sim 0.4$ and $k \sim 4$)
    - Lower values (of k or f) lead to more intervention

Causal Forecasting Models

- Assumes that demand is highly correlated with some environmental factors
- Model is built to relate the independent exogenous factors to the demand
- Examples:
  - Diapers ~ f(birth rates lagged by 1 year)
  - NFL Jerseys ~ f(team and individual performance)
  - New products ~ f(product lifecycle)
  - Promotional Items ~ f(marketing promotions & ads)
  - Regional Sales ~ f(household demographics in area)
  - Umbrellas / Fuel ~ f(weather, temperature, rain, etc.)
- Form of Dependent Variable dictates the method used
  - Continuous – takes any value
  - Discrete – takes only integer values
  - Binary – is equal to 0 or 1

OLS Linear Regression

- The relationship is described in terms of linear model
- The data (x_i, y_i) are the observed pairs from which we try to estimate the β coefficients to find the 'best fit'
- The error term, ε, is the 'unaccounted' or 'unexplained' portion
- The error terms are assumed to be iid ~N(0,σ) and catch all of the factors ignored or neglected in the model

$$y_i = \beta_0 + \beta_1 x_i$$

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad for\ i = 1, 2, \ldots n$$

Observed

Unknown

$$E(Y \mid x) = \beta_0 + \beta_1 x$$

$$StdDev(Y \mid x) = \sigma$$

OLS Linear Regression

- Residuals
  - Predicted or estimated values are found by using the regression coefficients, b.
  - Residuals, e_i, are the difference of actual – predicted values
  - Find the b's that "minimize the residuals"

$$\hat{y}_i = b_0 + b_1 x_i \quad for\ i = 1, 2, \ldots n$$

$$e_i = y_i - \hat{y}_i = y_i - b_0 + b_1 x_i \quad for\ i = 1, 2, \ldots n$$

- How should I measure the residuals?
  - Min sum of errors - shows bias, but not accurate
  - Min sum of absolute error - accurate & shows bias, but intractable
  - Min sum of squares of error – shows bias & is accurate

$$\sum_{i=1}^{n}\left(e_i^2\right) = \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{n}\left(y_i - b_0 - b_1 x_i\right)^2$$

The best model minimizes the residual sum of squares

OLS Linear Regression

- We can find the optimal values of b_0 and b_1 by taking first order conditions of the SSE:

$$\sum_{i=1}^{n}\left(e_i^{\,2}\right)=\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=\sum_{i=1}^{n}(y_i-b_0-b_1x_i)^2$$

- This gives us the following coefficients:

$$b_0=\bar{y}-b_1\bar{x}$$

$$b_1=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$$

OLS Linear Regression

- Expansion to multiple variables is straightforward
- So, for k variables we need to find k regression coefficients

$$Y_i = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_k x_{ki} + \varepsilon_i \quad for\ i = 1, 2, \ldots n$$

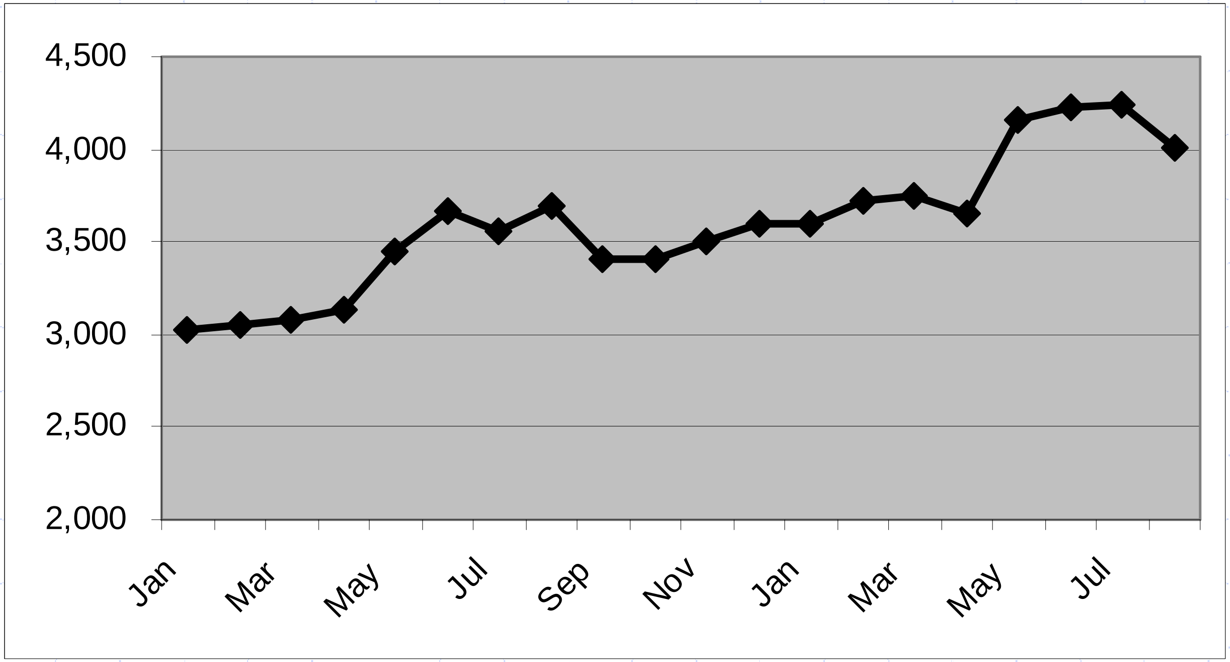
$$E(Y \mid x_1, x_2, \ldots, x_k) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

$$StdDev(Y \mid x_1, x_2, \ldots, x_k) = \sigma$$

$$\sum_{i=1}^{n}\left(e_i^{\,2}\right) = \sum_{i=1}^{n}\left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{n}\left(y_i - b_0 - b_1 x_{1i} - \ldots - b_k x_{ki}\right)^2$$

OLS Example

Month	Demand
Jan	3,025
Feb	3,047
Mar	3,079
Apr	3,136
May	3,454
Jun	3,661
Jul	3,554
Aug	3,692
Sep	3,407
Oct	3,410
Nov	3,499
Dec	3,598
Jan	3,596
Feb	3,721
Mar	3,745
Apr	3,650
May	4,157
Jun	4,221
Jul	4,238
Aug	4,008



- What do you see?

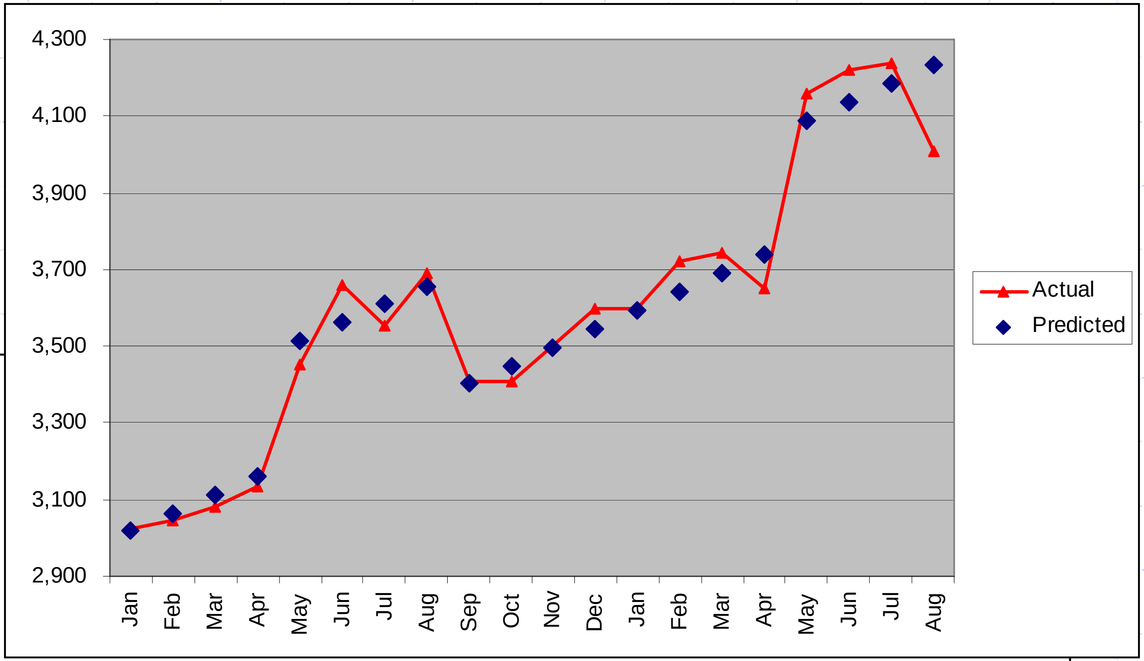
OLS Example

- Establish relationship
  - $F_i = f(X_{1i}, X_{2i}, \ldots X_{ni})$
  - $=\beta_0+\beta_1 X_{1i}+\beta_2 X_{2i}+\ldots+\beta_n X_{ni}$
  - F_i = Level + Trend + Season
    $= \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$

    Where X_{2i} = 1 if a summer month,
    = 0 o.w.
- Points to consider:
  - What if the trend is not linear?
  - How do I handle seasonality if it impacts the trend?
  - How does OLS treat old versus new data?
  - How much information do I need to keep on hand?

Month	Demand	Period	Summer
Jan	3,025	1	0
Feb	3,047	2	0
Mar	3,079	3	0
Apr	3,136	4	0
May	3,454	5	1
Jun	3,661	6	1
Jul	3,554	7	1
Aug	3,692	8	1
Sep	3,407	9	0
Oct	3,410	10	0
Nov	3,499	11	0
Dec	3,598	12	0
Jan	3,596	13	0
Feb	3,721	14	0
Mar	3,745	15	0
Apr	3,650	16	0
May	4,157	17	1
Jun	4,221	18	1
Jul	4,238	19	1
Aug	4,008	20	1

OLS Example (Excel)



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.979
R Square	0.958
Adjusted R Square	0.953
Standard Error	79.21
Observations	20

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	2442766.966	1221383.483	194.6730408	1.91955E-12
Residual	17	106658.4214	6274.024786		
Total	19	2549425.387			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	2,969.14	37.21	79.79	0.0000	2,890.62	3,047.65	2,890.62	3,047.65
Period	48.03	3.20	15.00	0.0000	41.27	54.79	41.27	54.79
Summer	303.51	37.70	8.05	0.0000	223.97	383.04	223.97	383.04

$$F_i = 2969 + 48\ (\text{Period}) + 304\ (\text{Summer\_Flag})$$

OLS Example (Excel)

Coefficient of Determination
$R^2 = 1\text{-}ESS/TSS = RSS/TSS$

Standard Error (estimate of σ around the regression line)

Sum of the Squares
Regression (RSS) $= \sum(\hat{y}-\bar{y})^2$
Error (ESS) $= \sum(y-\hat{y})^2$
Total (TSS) $= \sum(y-\bar{y})^2$

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.979
R Square	0.958
Adjusted R Square	0.953
Standard Error	79.21
Observations	20

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	2442766.966	1221383.483	194.6730408	1.91955E-12
Residual	17	106658.4214	6274.024786		
Total	19	2549425.387			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	2,969.14	37.21	79.79	0.0000	2,890.62	3,047.65
Period	48.03	3.20	15.00	0.0000	41.27	54.79
Summer	303.51	37.70	8.05	0.0000	223.97	383.04

95% Confidence Intervals

Regression Coefficients

Degrees of Freedom = n-k-1

Std Error of Regression Coeff (s_{bm})

t-Statistic (b_m/s_{bm})
Is b_m different from 0?
P-value tells you % conf.

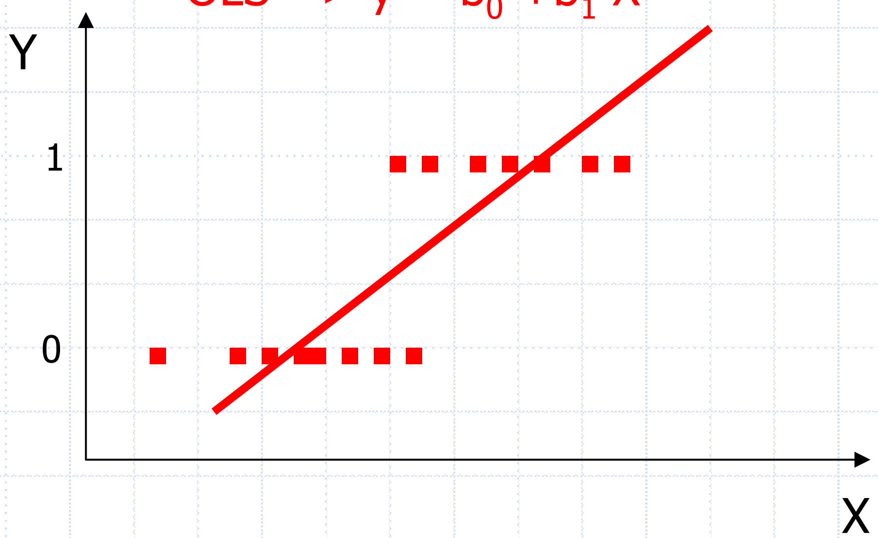
Coefficient of Determination (R^2)

- Measures Goodness of Fit of the model
- Captures the amount of variation that the model 'explains'
  - R^2=1-ESS/TSS = RSS/TSS
    - TSS = ESS + RSS
    - Variation of observed around mean = Variation of observed around estimated – Variation of estimated around the mean
- Generally, a higher R^2 is better, but . . .
  - Model needs to make sense
  - High R^2 does not indicate causality
  - It really depends on how the model is being used as to what is 'good enough'
  - The individual coefficients need to be tested

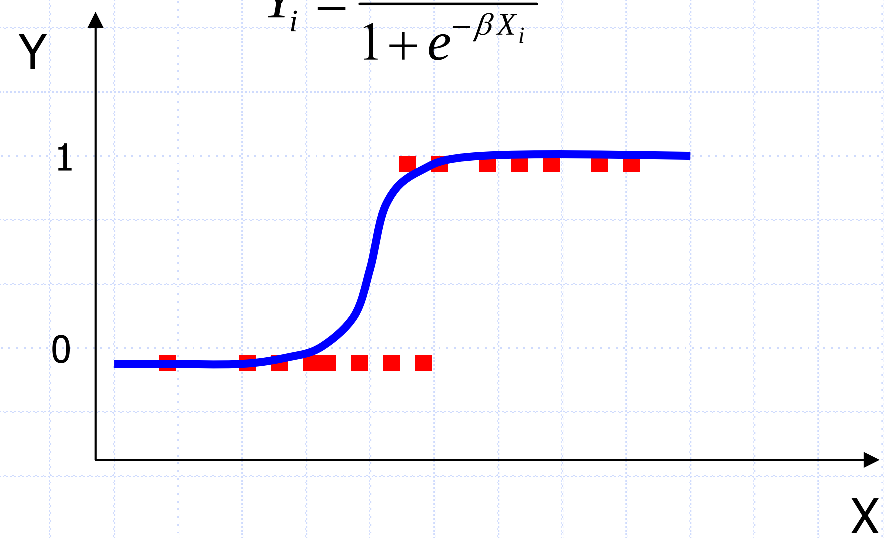
Discrete Choice Models

- What if you are predicting demand for one product over another?
  - Model Selections (Blue vs. Red Cars)
  - Mode Forecasting (pick one of many)

OLS => $y = b_0 + b_1 X$



$$Y_i = \frac{1}{1 + e^{-\beta X_i}}$$



Sales Forecasting Methods

- Expert Opinions
  - 44.8% Sales Force
  - 37.3% Executives
  - 14.9% Industry Surveys
- Statistical Models
  - 30.6% Naïve Model
  - 20.9% Moving Average
  - 11.2% Exp. Smoothing
  - 6.0% Regression
  - 3.7% Box-Jenkins

Source: Dalrymple (1987) Survey 134 companies

- Sales Forecast Errors (MAPE) by forecast horizons in years

Level	<.25 yrs	≤2 yrs	>2 yrs
Industry	8	11	15
Corporate	7	11	18
Product Group	10	15	20
Product Line	11	16	20
Product	16	21	26

Source: Mentzer & Cox (1984)

Misc. Forecasting Issues

- Data Issues
  - Sales data is not demand data
  - Transactions can aggregate and skew actual demand
  - Ordering quantities can dictate sourcing
  - Historical data might not exist
- Demand visibility can be skewed by level of echelon
  - Bullwhip effect
  - Collaborative Planning, Forecasting, and Replenishment (CPFR)
- Forecasting vs. Inventory Management
- Statistical Validity vs. Use and Cost of Model
- Demand is not always exogenous

Questions, Comments, Suggestions?

Massachusetts
Institute of
Technology