Demand Forecasting II
Causal Analysis

Chris Caplice
ESD.260/15.770/1.260 Logistics Systems
Sept 2006
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: 
\[
MD = \frac{\sum_{t=1}^{n} e_i}{n}
\]

3. Mean Absolute Deviation
\[
MAD = \frac{\sum_{t=1}^{n} |e_i|}{n}
\]

4. Mean Squared Error:
\[
MSE = \frac{\sum_{t=1}^{n} e_i^2}{n}
\]

5. Root Mean Squared Error:
\[
RMSE = \sqrt{\frac{\sum_{t=1}^{n} e_i^2}{n}}
\]

6. Mean Percent Error:
\[
MPE = \frac{\sum_{t=1}^{n} \frac{e_i}{D_t}}{n}
\]

7. Mean Absolute Percent Error:
\[
MAPE = \frac{\sum_{t=1}^{n} \frac{|e_i|}{D_t}}{n}
\]
Moving Average Forecasts

<table>
<thead>
<tr>
<th></th>
<th>MA3</th>
<th>MA10</th>
<th>MA20</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD</td>
<td>0.05</td>
<td>0.21</td>
<td>0.35</td>
</tr>
<tr>
<td>MAD</td>
<td>0.56</td>
<td>1.07</td>
<td>1.41</td>
</tr>
<tr>
<td>MSE</td>
<td>0.47</td>
<td>1.67</td>
<td>2.71</td>
</tr>
<tr>
<td>RMSE</td>
<td>0.68</td>
<td>1.29</td>
<td>1.65</td>
</tr>
<tr>
<td>MAPE</td>
<td>0.50%</td>
<td>0.96%</td>
<td>1.27%</td>
</tr>
</tbody>
</table>
Analysis of the Forecast

Are the forecast errors \( \sim N(0, \text{Var}(e)) \)?
- 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 \( \sim N(0, \text{Var}(e)) \)?
  - 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 $\sim 0$ if unbiased
- **Smoothed Error Tracking Signal** ($T_t$)
  - $T_t = \frac{z_t}{MAD_t}$
  - Where $z_t = \omega e_t + (1-\omega)z_{t-1}$ (smoothing constant)

**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 \(\beta\) coefficients to find the ‘best fit’
- The error term, \(\varepsilon\), is the ‘unaccounted’ or ‘unexplained’ portion
- The error terms are assumed to be iid \(\sim \text{N}(0, \sigma)\) and catch all of the factors ignored or neglected in the model

\[
y_i = \beta_0 + \beta_1 x_i
\]

\[
Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad \text{for } i = 1, 2, \ldots, n
\]

\[
E(Y \mid x) = \beta_0 + \beta_1 x
\]

\[
\text{StdDev}(Y \mid x) = \sigma
\]
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”

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

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

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

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

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:

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

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

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

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

\[
\text{StdDev}(Y \mid x_1, x_2, \ldots, x_k) = \sigma
\]

\[
\sum_{i=1}^{n} (e_i^2) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 x_{1i} - \ldots - b_k x_{ki})^2
\]
OLS Example

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

What do you see?
OLS Example

- **Establish relationship**
  
  \[ F_i = f(X_{1i}, X_{2i}, \ldots X_{ni}) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_n X_{ni} \]

  \[ F_i = \text{Level} + \text{Trend} + \text{Season} = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} \]

  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?

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

**SUMMARY OUTPUT**

<table>
<thead>
<tr>
<th>Regression Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
</tr>
<tr>
<td>R Square</td>
</tr>
<tr>
<td>Adjusted R Square</td>
</tr>
<tr>
<td>Standard Error</td>
</tr>
<tr>
<td>Observations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td>df</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>Regression 2</td>
</tr>
<tr>
<td>Residual 17</td>
</tr>
<tr>
<td>Total 19</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
<th>Lower 95.0%</th>
<th>Upper 95.0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2,969.14</td>
<td>37.21</td>
<td>79.79</td>
<td>0.0000</td>
<td>2,890.62</td>
<td>3,047.65</td>
<td>2,890.62</td>
</tr>
<tr>
<td>Period</td>
<td>48.03</td>
<td>3.20</td>
<td>15.00</td>
<td>0.0000</td>
<td>41.27</td>
<td>54.79</td>
<td>41.27</td>
</tr>
<tr>
<td>Summer</td>
<td>303.51</td>
<td>37.70</td>
<td>8.05</td>
<td>0.0000</td>
<td>223.97</td>
<td>383.04</td>
<td>223.97</td>
</tr>
</tbody>
</table>

\[ F_i = 2969 + 48 \text{ (Period)} + 304 \text{ (Summer\_Flag)} \]
# OLS Example (Excel)

## SUMMARY OUTPUT

**Regression Statistics**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
<td>0.979</td>
</tr>
<tr>
<td>R Square</td>
<td>0.958</td>
</tr>
<tr>
<td>Adjusted R Square</td>
<td>0.953</td>
</tr>
<tr>
<td>Standard Error</td>
<td>79.21</td>
</tr>
<tr>
<td>Observations</td>
<td>20</td>
</tr>
</tbody>
</table>

**ANOV**

<table>
<thead>
<tr>
<th></th>
<th>df</th>
<th>SS</th>
<th>MS</th>
<th>F</th>
<th>Significance F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression</td>
<td>2</td>
<td>2442766.966</td>
<td>1221383.483</td>
<td>194.673</td>
<td>1.91955E-12</td>
</tr>
<tr>
<td>Residual</td>
<td>17</td>
<td>106658.4214</td>
<td>6274.024786</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>19</td>
<td>2549425.387</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Coefficients**

<table>
<thead>
<tr>
<th></th>
<th>Coefficients</th>
<th>Standard Error</th>
<th>t Stat</th>
<th>P-value</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2,969.14</td>
<td>37.21</td>
<td>79.79</td>
<td>0.0000</td>
<td>2,890.62</td>
<td>3,047.65</td>
</tr>
<tr>
<td>Period</td>
<td>48.03</td>
<td>3.20</td>
<td>15.00</td>
<td>0.0000</td>
<td>41.27</td>
<td>54.79</td>
</tr>
<tr>
<td>Summer</td>
<td>303.51</td>
<td>37.70</td>
<td>8.05</td>
<td>0.0000</td>
<td>223.97</td>
<td>383.04</td>
</tr>
</tbody>
</table>

---

### Coefficient of Determination

\[ R^2 = 1 - \frac{\text{ESS}}{\text{TSS}} = \frac{\text{RSS}}{\text{TSS}} \]

### Standard Error (estimate of \( \sigma \) 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)**: \[ \sum (y - \bar{y})^2 \]

### Degrees of Freedom

\[ \text{Degrees of Freedom} = n - k - 1 \]

### 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’
  - $R^2 = 1 - \frac{ESS}{TSS} = \frac{RSS}{TSS}$
    - $TSS = ESS + RSS$
    - 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)

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

Sales Forecast Errors (MAPE) by forecast horizons in years

<table>
<thead>
<tr>
<th>Level</th>
<th>&lt;.25 yrs</th>
<th>≤2 yrs</th>
<th>&gt;2 yrs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Industry</td>
<td>8</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>Corporate</td>
<td>7</td>
<td>11</td>
<td>18</td>
</tr>
<tr>
<td>Product Group</td>
<td>10</td>
<td>15</td>
<td>20</td>
</tr>
<tr>
<td>Product Line</td>
<td>11</td>
<td>16</td>
<td>20</td>
</tr>
<tr>
<td>Product</td>
<td>16</td>
<td>21</td>
<td>26</td>
</tr>
</tbody>
</table>

Source: Dalrymple (1987) Survey 134 companies

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?