Airport Forecasting

Prof. Richard de Neufville

Istanbul Technical University
Air Transportation Management
M.Sc. Program

Airport Planning and Management
Module 07
January 2017
Forecasting In Practice

• **Objective:** To present procedure.

• **Topics:**
  1. Premises
  2. Many Assumptions underlie forecast methods
  3. Basic mechanics of forecast methods
  4. Principles for Practice
  5. Recommended Procedure
  6. Mexico City Example
  7. Current International Considerations
  8. Summary
Premises

• Forecasting is an Art, not a Science -- too many assumptions
  not a statistical exercise -- too many solutions

• Forecasts are Inherently Risky
Results of a study of TAF

Errors in 5 year TAF

Note: Average error ~ 11%

Adapted from: Terminal Area Forecast (TAF) Accuracy Assessment Results
Assumptions behind any forecasting exercise

- **Span of data -- number of periods or situations (10 years? 20? 30?)**

**Domestic RPMs Projection vs. Reality**

Past 10 years: almost level...

Past 20 years: Strong rise…
Assumptions behind any forecasting exercise

- Span of data -- number of periods or situations (10 years? 20? 30?)
- Variables -- which ones in formula (price? income? employment? etc)
- Form of variables -- total price? price relative to air? To ground?
- Form of equation -- linear? log-linear? translog? Logit?

Logical House of Cards
Choice of variables

• Note first: The more variables you include, the better the statistics in model, the better the fit!

• Why is that?

• Because procedures for creating statistical model only include variables to extent they improve $R^2$
Common forms of forecasting equations

• **Linear**
  \[ \text{Pax} = \text{Population} \,[a + b(\text{Income})+c(\text{Yield})…] \]

• **Exponential**
  \[ \text{Pax} = \{a \,[\text{Yield}]^b\}{c \,[\text{population}]^d} \{\text{etc}…\} \]

• **Exponential in Time**
  \[ \text{Pax} = a \,[e]^rt \]
  where \( r \) = rate per period
  and \( t \) = number of periods

• **Benefits of each?**
Fundamental Mathematics of Regression Analysis

- Linear equations
  - Logarithm of exponential form => linear

- Define “fit”
  - = sum of squared differences of equation and data, \( \Sigma (y_1 - y_2)^2 \)
  - => absolute terms, bell-shaped distribution

- Optimize fit
  - differentiate fit, solve for parameters
  - R-squared measures fit (0 < R^2 < 1.0)
Let’s discuss meaning of correlation for a moment

• There is well-established good correlation between:
  (Damage at Fire) and (Number of Firemen)

• What do I conclude about how Firemen cause damage?
  ➔ Should I send less firemen to fire?

• The correlation is “spurious”:
  Big fires => damage, firemen sent
Ambiguity of Results: Many ‘good’ results possible

- Common variables (employment, population, income, etc) usually grow exponentially \( \sim a(e)^{rt} \)
- They are thus direct functions of each other
  \[ a(e)^{rt} = [(a/b)(e)^{(r/p)t}]b(e)^{pt} \]
- Easy to get ‘good’ fit
  \[ \rightarrow \text{See Miami example} \]
Forecasts of International Passengers (Millions/Year) for Miami/International

<table>
<thead>
<tr>
<th>Forecast</th>
<th>Method</th>
<th>Case</th>
<th>Forecast 2020</th>
<th>Actual 1990</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>Dade Co.</td>
<td>16.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dade/Broward</td>
<td>16.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dade/Broward (Non-Linear)</td>
<td>21.89</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yield and Per Capita Personal Income</td>
<td>Dade Co.</td>
<td>19.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dade/Broward</td>
<td>22.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dade/Broward (Non-Linear)</td>
<td>20.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Time Series</td>
<td>Dade Co.</td>
<td>19.84</td>
<td></td>
<td>10.01</td>
</tr>
<tr>
<td></td>
<td>Dade/Broward</td>
<td>20.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dade/Broward (Non-Linear)</td>
<td>57.61</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Per Capita Personal Income</td>
<td>Dade Co.</td>
<td>28.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dade/Broward</td>
<td>25.57</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dade/Broward (Non-Linear)</td>
<td>53.79</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Share (US Int’l Pax)</td>
<td>Maximum</td>
<td>57.61</td>
<td>576 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>27.49</td>
<td>275 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td>21.20</td>
<td>212 %</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>16.60</td>
<td>166 %</td>
<td></td>
</tr>
<tr>
<td>Share (US Reg’l Rev.)</td>
<td>Preferred</td>
<td>37.76</td>
<td>377 %</td>
<td></td>
</tr>
</tbody>
</table>

Source: Landrum and Brown (Feb. 5, 1992)
## Forecasts of Domestic Passengers (Millions per year) for Miami/International

<table>
<thead>
<tr>
<th>Forecast</th>
<th>Forecast Method</th>
<th>Case</th>
<th>Forecast 2020</th>
<th>Actual 1990</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>Dade Co.</td>
<td></td>
<td>13.96</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dade/Broward</td>
<td></td>
<td>15.35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dade/Broward (Non-Linear)</td>
<td></td>
<td>17.74</td>
<td></td>
</tr>
<tr>
<td>Yield and Per Capita Personal Income</td>
<td>Dade Co.</td>
<td></td>
<td>19.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dade/Broward</td>
<td></td>
<td>19.69</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dade/Broward (Non-Linear)</td>
<td></td>
<td>19.13</td>
<td></td>
</tr>
<tr>
<td>Time Series</td>
<td>Dade Co.</td>
<td></td>
<td>17.41</td>
<td>9.92</td>
</tr>
<tr>
<td></td>
<td>Dade/Broward</td>
<td></td>
<td>18.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dade/Broward (Non-Linear)</td>
<td></td>
<td>40.05</td>
<td></td>
</tr>
<tr>
<td>Per Capita Personal Income</td>
<td>Dade Co.</td>
<td></td>
<td>26.58</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dade/Broward</td>
<td></td>
<td>24.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dade/Broward (Non-Linear)</td>
<td></td>
<td>42.40</td>
<td></td>
</tr>
<tr>
<td>Share of US Traffic</td>
<td>Maximum</td>
<td></td>
<td>42.40</td>
<td>427 %</td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td></td>
<td>22.97</td>
<td>232 %</td>
</tr>
<tr>
<td></td>
<td>Median</td>
<td></td>
<td>19.69</td>
<td>198 %</td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td></td>
<td>13.96</td>
<td>141 %</td>
</tr>
<tr>
<td></td>
<td>Preferred</td>
<td></td>
<td>15.35</td>
<td>155 %</td>
</tr>
</tbody>
</table>

Source: Landrum and Brown (Feb. 5, 1992)
Note Use of “preferred” forecast

- Forecasts obtained statistically often “don’t make sense”
- Forecasters often disregard statistical results (expensive, misleading), substituting intuition (cheap)
- E.g.: NE Systems Study (SH&E, 2005) “The long-term forecast growth… was inconsistent with… expectations…[and] were revised to… more reasonable levels”
# Domestic Pax for Miami update for 2010, 2012

<table>
<thead>
<tr>
<th>Forecast Method and Variant</th>
<th>Forecast Method and Variant</th>
<th>Forecast</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Method</strong></td>
<td><strong>Population</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Used (form)</td>
<td>Dade Country</td>
<td>13.96</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dade and Broward</td>
<td>15.35</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dade and Broward (non-linear)</td>
<td>17.74</td>
<td></td>
</tr>
<tr>
<td><strong>Yield and Per Capita Personal Income</strong></td>
<td>Dade County</td>
<td>19.87</td>
<td>9.92</td>
</tr>
<tr>
<td></td>
<td>Dade and Broward</td>
<td>19.69</td>
<td>17.4</td>
</tr>
<tr>
<td></td>
<td>Dade and Broward (non-linear)</td>
<td>19.13</td>
<td></td>
</tr>
<tr>
<td><strong>Time Series</strong></td>
<td>Dade County</td>
<td>17.41</td>
<td>18.8</td>
</tr>
<tr>
<td></td>
<td>Dade and Broward</td>
<td>18.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Dade and Broward (non-linear)</td>
<td>40.05</td>
<td></td>
</tr>
<tr>
<td><strong>Per Capita Personal Income</strong></td>
<td>Dade County</td>
<td>26.58</td>
<td>9.92</td>
</tr>
<tr>
<td></td>
<td>Dade and Broward</td>
<td>24.34</td>
<td>17.4</td>
</tr>
<tr>
<td></td>
<td>Dade and Broward (non-linear)</td>
<td>42.40</td>
<td></td>
</tr>
<tr>
<td><strong>Share of US</strong></td>
<td>Maximum</td>
<td>42.40</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Average</td>
<td>22.97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Medium</td>
<td>19.69</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Minimum</td>
<td>13.96</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>Preferred</strong></td>
<td>15.35</td>
<td></td>
</tr>
</tbody>
</table>
“Miami set a new all-time record for annual passenger traffic in 2011 with 35.7 million passengers”

BUT:

“The previous record was set in 1997 when the airport welcomed 34.5 million passengers.”

Principles for forecasting in practice

- Detailed Examination of Data
  Statistics are often inconsistent, wrong, or otherwise inappropriate for extrapolation

- Extrapolation for Short Term,
  About five years

- Scenarios for Long Term,
  Allowing for basic changes

- Ranges on Forecasts,
  As wide as experience demonstrates
Recommended Procedure

1. **Examine Data**
   compare sources, check consistency

2. **Identify Possible Causal Factors**
   relevant to site, period, activity

3. **Do regression, extrapolate for short term,**
   apply historical ranges on forecasts

4. **Identify future scenarios**

5. **Project ranges of possible consequences**

6. **Validate Plausibility**
   compare with elsewhere
Passengers, Mexico City International Airport (AICM)
Mexico City -- Data Problems

- **Typographical Error**
  Seen by examination of primary data

- **Double Counting**
  Introduced by a new category of data

- **New Definitions of Categories**
  Detected by anomalies in airline performance (pax per aircraft) for national, internat’l traffic

These problems occur anywhere
Passengers Through AICM (Corrected Version)

Corrected Air Passengers Through Mexico City ($10^6$)

- Total
- National
- International

1960 1968 1976
Mexico City
Causes of Trends

- Economic Boom
  Post 1973 oil prosperity

- Recession Elsewhere
  Affecting international traffic

- Population Growth

- Fare Cuts
  Relative to other commodities
Population Increase of Mexico City’s Metro Area

The 2015 population of metropolitan Mexico City was about 21.2 million!
Trend of International Air Fares (at Constant Prices)
Mexico City -- Note

- Traffic formula based on these variables (or others) does not solve forecasting problem.

- Why?

- Formula displaces problem, from traffic to other variables.

- How do we forecast values of other variables (population, etc)?
Short-Range Forecasts, National Passengers, AICM

Forecast National Passengers for Mexico City (millions)

Actual 2010 Was 15.6
Short-Range Forecasts, International Pax. AICM

Forecast International Passengers for Mexico City (millions)

Actual 2010 was 8.5

Forecast High
Medium
Low

Corrected Series
Mexico City -- Elements of Long-range Scenarios

- **Demographics**
  - Rate of Population Increase
  - Relative Size of Metropolis

- **Economic Future**

- **Fuel Prices and General Costs**

- **Technological, Operational Changes**

- **Timing of Saturation**
Long-range Scenarios

• **New Markets**
  - Japan, Pacific Rim, United Europe

• **More Competition**
  - Deregulation, Privatization
  - Transnational Airlines

• **New Traffic Patterns**
  - Direct flights bypassing Mexico City
  - More Hubs (Bangkok, Seoul?)
  - New Routes, such as over Russia
Long Term AICM Forecasts, validated by data elsewhere

- Mexico City Forecast (High)
- Mexico City Forecast (Mid)
- Mexico City Forecast (Low)
- Los Angeles
- London
- Osaka

Actual:
- 2010: 24.1 M
- 2015: 38.4 M

In 2015
Summary

• **Forecasting is not a Science**
  → too many assumptions
  → too much ambiguity

• **Regression analysis for short term**
  → Apply historical ranges on projections

• **Scenarios for Long range**
  → compare with experience elsewhere

• **STRESS UNCERTAINTY**