REVENUE FORECASTING

Not Your Dad’s Magic Eight Ball

Prepared for the NCSL Fiscal Analysts Seminar, October 21, 2014

Jim Landers, Office of Fiscal and Management Analysis, Indiana Legislative Services Agency
Forecasts

- Extrapolate from past actual values
- Project future values
- Are not deterministic

Our discussion this morning

- State Revenue Forecasting Processes
- Forecasting methods
- Conducting a sales tax forecast
STATE REVENUE FORECASTING PROCESSES

Executive & Legislative Processes

- Executive branch forecast
  - Forecast by single executive branch agency
- Dueling forecasts
  - Executive branch and legislative agencies

Consensus Processes

- 27 states
- Collaborative process
- Includes executive and legislative branch representatives
- Research
  - Forecast errors tend to be reduced
OVERVIEW OF INDIANA’S PROCESS
FORECASTING METHODS

Naive
- Expert
- Judgmental

Incremental
- Extrapolative
- Trend

Causal
- Multivariate
- Econometric
A TIME SERIES

- Is a set of single data points
- Is recorded sequentially over time
- Has 3 components
  - Trend
  - Seasonal
  - Cyclical
- Has historical patterns that:
  - could continue into the future
  - we can explain and use to forecast
ANYBODY KNOW WHAT THESE ARE?

Univariate Methods

Multivariate Methods
UNIVARIATE FORECASTING METHODS

Requires a time series for a single variable

Forecast variable could be tax revenue or tax base
  • Account for tax rate changes
  • Account for significant tax base changes

Assumes historical patterns or regularities can be explained with one variable

Autoregressive (AR) models
  • Forecast future values based on combination of prior values of the variable

Moving average (MA) models
  • Forecast future values based on a moving average of prior forecast errors
MULTIVARIATE FORECASTING METHODS

Econometric, multiple regression models

Requires a time series for multiple variables

- Forecast variable could be tax revenue or tax base variable
- Predictors such as economic and policy variables

Assumes that historical patterns or regularities can be explained by other correlated variables
What's a Sales Tax?
INDIANA SALES TAX

Tax on retail final consumption sales

Current rate is 7%.
• Rate changes in 1983 (to 5%), 2003 (to 6%), and 2008 (to 7%)

Applies to sales of tangible property
• Durable goods (autos, appliances, furniture, etc.)
• Nondurable goods (clothing, household goods, etc.)
• Food for consumption at home is exempt from tax
• Prescription drugs are exempt from tax

Applies to limited number of services

Applies to some intermediate business purchases
• Maybe 33% of tax revenue
SALES TAX REVENUE SERIES

Sales Tax Revenue vs Linear (Sales Tax Revenue)
FORECASTING CAVEAT

High Octane Models

Low Octane Data
SOME DATA PROBLEMS

- Series length
  - 1970s, 1980s.
  - Spending on services
  - Recessions

- Adjustments for base changes
  - Based on estimates

- Auto vs. other sales
What variables explain variation in or are correlated with annual sales tax revenue?
POTENTIAL EXPLANATORY VARIABLES

Policy
- Tax rate changes
- Tax base changes

Consumer
- Personal income
- Personal income less transfer payments
- Wages and salaries

Business
- GDP
- % Change in GDP

Other
- Personal Savings Rate
- Personal Consumption Expenditures
- Unemployment Rate
SALES TAX REVENUE SERIES
MODEL VARIABLES

Predicted
- Sales Tax Revenue (in millions)

Predictors
- Pers. Income (in millions)
- Sales Tax Rate
- % Change in GDP

(1) \[ S = a + b \times I \]
(2) \[ S = a + b \times I + c \times T \]
(3) \[ S = a + b \times I + c \times T + d \times G \]

- S=Sales Tax Revenue
- I=Personal Income
- T=Sales Tax Rate
- G=% Change in GDP
# Reading the Model Results?

### SUMMARY OUTPUT

<table>
<thead>
<tr>
<th>Regression Statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Multiple R</td>
<td>0.998793735</td>
</tr>
<tr>
<td>R Square</td>
<td>0.997588926</td>
</tr>
<tr>
<td>Adjusted R Square</td>
<td>0.99727265</td>
</tr>
<tr>
<td>Standard Error</td>
<td>80.65072331</td>
</tr>
<tr>
<td>Observations</td>
<td>24</td>
</tr>
</tbody>
</table>

### ANOVA

<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>3</td>
<td>53825440.5</td>
<td>17941813.5</td>
<td>2758.352749</td>
<td>2.45384E-26</td>
</tr>
<tr>
<td>Residual</td>
<td>20</td>
<td>130090.7834</td>
<td>6504.53917</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>23</td>
<td>53955531.28</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>-2837.963545</td>
<td>166.1603826</td>
<td>-17.07966424</td>
<td>2.14673E-13</td>
<td>-3184.368029</td>
<td>-2911.35906</td>
</tr>
<tr>
<td>Sales Tax Rate</td>
<td>57322.05267</td>
<td>4340.648985</td>
<td>13.20587149</td>
<td>2.45915E-11</td>
<td>48267.61755</td>
<td>66376.48779</td>
</tr>
<tr>
<td>Personal Income (millions)</td>
<td>0.022414086</td>
<td>0.000735823</td>
<td>30.46128942</td>
<td>3.11124E-18</td>
<td>0.020879187</td>
<td>0.023948986</td>
</tr>
<tr>
<td>Chng. GDP</td>
<td>-767.9184545</td>
<td>798.9933116</td>
<td>-0.961107488</td>
<td>0.347974793</td>
<td>-2434.589297</td>
<td>898.752388</td>
</tr>
</tbody>
</table>
READING THE ESTIMATION RESULTS

- Model Explanatory Power
  - Model Significance
  - R-Square

- Predictor Impacts
  - Regression Coefficients
  - Coefficient Significance

- Other
  - Elasticities
  - Std. Error of Estimate
# Reading the Estimation Results

<table>
<thead>
<tr>
<th>Variable</th>
<th>Model 1*</th>
<th>Model 2*</th>
<th>Model 3*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>-1103.31**</td>
<td>-2909.00**</td>
<td>-2837.96**</td>
</tr>
<tr>
<td>$t, p$-value</td>
<td>-6.01, 0.00</td>
<td>-19.58, 0.00</td>
<td>-17.07, 0.00</td>
</tr>
<tr>
<td>Personal Income (in millions)</td>
<td>0.031** (e=1.26)</td>
<td>0.023** (e=.917)</td>
<td>0.022** (e=.912)</td>
</tr>
<tr>
<td>$t, p$-value</td>
<td>29.99, 0.00</td>
<td>31.07, 0.00</td>
<td>30.46, 0.00</td>
</tr>
<tr>
<td>Tax Rate</td>
<td>-</td>
<td>57617.69** (e=0.777)</td>
<td>57322.05** (e=0.773)</td>
</tr>
<tr>
<td>$t, p$-value</td>
<td>-</td>
<td>13.33, 0.00</td>
<td>13.20, 0.00</td>
</tr>
<tr>
<td>% Change GDP</td>
<td>-</td>
<td>-</td>
<td>-767.92</td>
</tr>
<tr>
<td>$t, p$-value</td>
<td>-</td>
<td>-</td>
<td>-0.96, 0.35</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.975**</td>
<td>.997**</td>
<td>.997**</td>
</tr>
<tr>
<td>$F, p$-value</td>
<td>899.61, 0.00</td>
<td>4152.14, 0.00</td>
<td>2758.35, 0.00</td>
</tr>
<tr>
<td>Standard Error</td>
<td>241.95</td>
<td>80.50</td>
<td>80.65</td>
</tr>
</tbody>
</table>

* $n=24$ **Statistically significant at .01.**
PREDICTION ERROR

\[ E = A_t - P_t \]

- \( P_t \) = Predicted value in period \( t \)
- \( A_t \) = Actual value in period \( t \)
- \( t \) = periods 1 to \( t \)

- Vertical deviations between actual values (white dots) and predicted values (blue regression line)
Measures of both precision and bias

- Measures based on deviation of predicted values from actual values
- Measures based on the mean of these deviations
- Relative (percentage) measures better for comparison

Bias measurement

- Mean Percentage Error (MPE)

Precision measurement

- Mean Absolute Percentage Error (MAPE)
Effect Measures

**MPE**

\[
MPE = \frac{1}{t} \sum_{t=1}^{t} \frac{A_t - P_t}{A_t}
\]

- \( P_t \) = Predicted value in period \( t \)
- \( A_t \) = Actual value in period \( t \)
- \( t \) = periods 1 to \( t \)

- Indicates the average percentage difference between:
  1. values predicted by the model
  2. actual values used to estimate the model

- It tells us:
  1. whether the model typically overestimates or underestimates
  2. the extent of the overestimation or underestimation
MAPE

\[
MAPE = \frac{\sum_{t=1}^{t} |A_t - P_t|}{A_t}
\]

\(| | = \text{absolute value}

- \(P_t=\text{predicted value in period } t\)
- \(A_t=\text{actual value in period } t\)
- \(t=\text{periods 1 to } t\)

MAPE

- Indicates the average percentage difference between:
  1. values predicted by the model
  2. actual values used to estimate the model
  3. in absolute value terms

- It tells us on average by how much the model predictions miss actuals over time
EX POST FORECAST

Checking how the model might forecast

In-sample/Out-of-sample analysis

If our forecast horizon is 2 years into the future we should at least do a 2 year ex post forecast

Re-estimate the model using the time series but leave out the last 2 years

Use the re-estimated model to forecast the last 2 years of the series

Measure the ex post forecast error – forecast vs. actual
EX ANTE ERROR ANALYSIS

Starting point for next forecast
- Test occurs after actuals have come in for a forecast period
- Measures model performance
- Measures performance of the separate forecast of the predictors specified in the model

Measures forecast error
- The difference between actual and forecast values

Measures the “Model Error”
- Share of forecast error attributable to the model specification

Measures the “Variable Error”
- Share of forecast error attributable to the forecast of the predictors specified in the model
EX ANTE ERROR ANALYSIS

Step 1 – measure the forecast error
  • Subtract the forecast value from the actual value
  • Negative value means over-forecast
  • Positive value means under-forecast

Step 2 – simulate forecast
  • Generate a “simulated” forecast with the forecast model and actual predictor values

Step 3 – measure the model error
  • Subtract the simulated forecast value from the actual value

Step 4 - measure the variable error
  • Subtract the model error from the forecast error