Ed Leamer’s Forecasting Cookbook

To make a delicious model, follow the following steps, but be sure to do some tasting along the way. Add sugar or salt as necessary.

1) **LIST CATEGORIES OF DRIVERS**: Select the variable to be forecast and identify at least five categories of drivers. Cast your net wide.

   For example some categories of drivers that might be used for forecasting sales of a product are: Customer income level, product price, price of complements (e.g. gas is a complement of autos), price of substitutes/competitors, product quality, consumer expectations.

   The categories of drivers should be different for short-term and for long-term forecasts. Excess capacity is a short-run driver of GDP growth. Capacity expansion is a long-run driver. Do you understand the difference?

2) **IMAGINE MEASUREMENTS: OF THE DRIVERS**. For each of the categories identify at least two ways it might be measured.

   For example, expectations might have something to do with the unemployment level (can't be very optimistic if you are unemployed) or about the recent growth in the economy (if things are going well, they probably will continue).

   Do not use the leading indicator as a predictor because it is an index (a mixture of a bunch of different series) and cannot be linked clearly to any category of drivers.

3) **MAKE ACTUAL MEASUREMENTS**: Select the three most important drivers and be sure that you can collect relevant data. If you cannot measure important drivers, make a note of the problem when you discuss results.

4) **IDENTIFY THE SALIENT FEATURES OF THE VARIABLE(S) OF INTEREST**

   Make a graph of the forecast variable(s) over time. Looking at the graph, identify at least two salient features that need to be understood. Compare with driver variables.

5) **TRANSFORM TO STATIONARY VARIABLES**. Define all the variables in a way that makes them approximately stationary.

   Use growth rates. (e.g. the growth of GDP) All growth rates should be annualized compound rates of growth. If the data y(t) are annual, the growth rate is

   \[ gy(t) = 100 \times \left( \frac{y(t)}{y(t-1)} \right) - 100 \]
If the data \( y(t) \) are quarterly, the growth rate is
\[
gy(t) = 100 \times \left[ \frac{y(t)}{y(t-n)} \right]^4 - 100
\]
If the data stretch over \( n \) years, the growth rate is
\[
gy(t) = 100 \times \left[ \frac{y(t)}{y(t-n)} \right]^{(1/n)} - 100
\]

Use proportions (e.g. the proportion of the labor force that is unemployed, savings as a percentage of GDP)

6) **FORM BENCHMARK SIMPLE EXTRAPOLATIVE FORECASTS OF THE STATIONARY VARIABLE(S)**  
An autoregressive model will do the job.

7) **MAKE SCATTER DIAGRAMS AND LINE DIAGRAMS COMPARING THE (STATIONARY) VARIABLE OF INTEREST WITH THE MEASURED DRIVERS.**

8) **BUILD A MODEL; HAVE AN ATTITUDE**

    **USE SEVERAL LAGGED DEPENDENT VARIABLES**
    If the growth rate of \( y \), \( gy(t) \), is the variable that is forecasted, include either \( gy(t-1) \) or the acceleration variable \( gy(t-1) - gy(t-2) \).
    If the percentage \( p(t) \) is the forecast variable, include either \( p(t-1) \) or the change variable \( p(t-1) - p(t-2) \).
    The data can help decide which is better.

    **INCLUDE IN THE MODEL AT MOST THREE DRIVER VARIABLES. USE CONCURRENT VARIABLES AND LAGGED VARIABLES:**
    If \( gx \) is a driver then include \( gx(t) \), or \( gx(t) - gx(t-1) \). The data can help decide which is the better driver.

    **HAVE AN ATTITUDE. LET THE DATA SPEAK IF THEY CAN, BUT OTHERWISE DON’T BUDGE AN INCH - YOU SELECT THE COEFFICIENTS**

    **EXCLUDE DOUBTFUL DRIVERS WITH t-values LESS THAN ONE in ABSOLUTE VALUE**
    This is a very dangerous practice, but do it anyway.

9) **CHECK TO MAKE SURE YOUR MODEL EXPLAINS THE SALIENT FEATURES IDENTIFIED IN STEP 4.**
    If necessary, translate back to the untransformed series.
10) **TELL A STORY.**

   The story is the most important step. Most of your customers want to buy the story and could care less about the forecast.

11) **BUILD ANOTHER MODEL TO EXTRAPOLATE DRIVERS THAT ARE SENSIBLY EXTRAPOLATED FROM HISTORICAL DATA** (e.g. interest rates)

12) **FORM ALTERNATIVE SCENARIOS FOR DRIVERS THAT ARE NOT SENSIBLY EXTROPLATED FROM HISTORICAL DATA** (e.g. Medicaid)
EXAMPLE: FORECASTING REAL GDP

1) LIST CATEGORIES OF DRIVERS:

   MONETARY POLICY
   FISCAL POLICY
   CAPACITY
   EXCESS CAPACITY

2) IMAGINE MEASUREMENTS: OF THE DRIVERS.

   MONETARY POLICY: Interest rates
   FISCAL POLICY: Defense expenditures
   CAPACITY: Historical investment
   EXCESS CAPACITY: Unemployment rate

3) MAKE ACTUAL MEASUREMENTS:

4) IDENTIFY THE SALIENT FEATURES OF THE VARIABLE(S) OF INTEREST

![Graph showing GDP growth with logarithmic scale]

The logarithmic scale allows us to see the slowdown in growth of GDP. There are at least four distinct dips - recessions.
It looks like the Vietnam cutback in defense is associated with a GDP slowdown in the early 1970s and the Reagan defense build-up is associated with the Reagan economic expansion in the 1980s. Then again the Cold War cutbacks are suspiciously timed to the recession in the early 1990s.

5) TRANSFORM TO STATIONARY VARIABLES.
6) **FORM BENCHMARK SIMPLE EXTRAPOLATIVE FORECASTS OF THE STATIONARY VARIABLE(S)**  
An autoregressive model will do the job.

```
<table>
<thead>
<tr>
<th>Forecast and Confidence Bands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth Rate</td>
</tr>
<tr>
<td>Lower Bound</td>
</tr>
<tr>
<td>Upper Bound</td>
</tr>
</tbody>
</table>
```

7) **MAKE SCATTER DIAGRAMS AND LINE DIAGRAMS COMPARING THE (STATIONARY) VARIABLE OF INTEREST WITH THE MEASUR ED DRIVERS.**

```
<table>
<thead>
<tr>
<th>Smoothed GDP Growth and Smoothed Defense Growth</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDEFSM</td>
</tr>
<tr>
<td>GSM</td>
</tr>
</tbody>
</table>
```

This picture doesn’t show much.
Not much apparent association, whether smoothed data or not.

8) **BUILD A MODEL; HAVE AN ATTITUDE**

<table>
<thead>
<tr>
<th>LS // Dependent Variable is G</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample(adjusted): 1959:4 1995:4</td>
</tr>
<tr>
<td>Included observations: 145 after adjusting endpoints</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Std. Error</th>
<th>t-Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>2.049807</td>
<td>0.448538</td>
<td>4.569975</td>
</tr>
<tr>
<td>G(-1)</td>
<td>0.262496</td>
<td>0.083027</td>
<td>3.161556</td>
</tr>
<tr>
<td>G(-2)</td>
<td>0.097401</td>
<td>0.081714</td>
<td>1.191985</td>
</tr>
<tr>
<td>GDEF-GDEF(-1)</td>
<td>0.048934</td>
<td>0.032963</td>
<td>1.484534</td>
</tr>
</tbody>
</table>

| R-squared | Mean dependent var | 3.198293 |
| Adjusted R-squared | S.D. dependent var | 3.944108 |
| S.E. of regression | 3.777180 |

This model allows for abrupt changes in defense to cause economic slowdowns.
9) **CHECK TO MAKE SURE YOUR MODEL EXPLAINS THE SALIENT FEATURES IDENTIFIED IN STEP 4.**

\[
predicted_G = 2.052942 + 0.25876688*G(-1) + 0.096443468*G(-2) + 0.06307036*GDEF - 0.034289602*GDEF(-1) \\
predicted_gdp = predicted_gdp(-1)*(1+predicted_g/100)^{(1/4)}
\]

This model takes as given the actual lagged values of G and GDEF and forms predictions within the sample period. This allows recessions/slow growth periods to be explained by either defense cutbacks or a slow recent history. These predicted growth rates are translated into levels of GDP, starting at the actual value in 1959:4. \(predicted_gdp(1959:4) = gdp(1959:4)\).

Note the model doesn’t capture the slow-down. It captures a little bit of the dips, but not much. I am not too happy with this. I am thinking of adding some more variables. What about \(POST68 = 1\) if \(t>1968\) and 0 otherwise to capture the slowdown? Or maybe a time trend?

10) **TELL A STORY.**

GDP is hard to predict. The defense expenditures appear to have some impact on GDP growth, but not much, probably because defense is too small a share of GDP and generally pretty predictable. But sharp swings in defense expenditures can make a difference.

11) **BUILD ANOTHER MODEL TO EXTRAPOLATE DRIVERS THAT ARE SENSIBLY EXTRAPOLATED FROM HISTORICAL DATA** (e.g. interest rates)
12) FORM ALTERNATIVE SCENARIOS FOR DRIVERS THAT ARE NOT SENSIBLY EXTROPLATED FROM HISTORICAL DATA (e.g. Medicaid)