Class 7. Prediction, Transformation and Multiple Regression.

1 Today's material

Prediction

Transformation

Multiple regression

- Robust regression
- Bootstrap

2 Prediction

Two types corresponding to the "data = signal + noise" paradigm.

Prediction of just the signal or prediction that also includes the noise.

Prediction in the range of the data (interpolation) is pretty safe.

Prediction out of the range of the data (extrapolation) is extremely dangerous.

Prediction for a new observation has 3 sources of uncertainty.

- The fit is not quite right – uncertainty in the true regression line.
- There's variability about the regression line – noise.
- There is uncertainty because this may not be the correct model – model misspecification.

3 Transformation

3.1 Why transform?

Upside: make life easy both practically (problems may evaporate, e.g. outliers become less severe) and theoretically (normal theory results, t-tests, p-values are credible)

Downside: may be hard to interpret

Rationale:

Symmetry – "middle" well defined

Easier to compare with normal (ie heavy tailed).

Methodology may require symmetry (ie normal theory)

Facilitates comparisons between observations that are on the same scale but far apart, (ie changes in Microsoft sales and changes in Apple's).

May be more interpretable – aid in decision making. Unit costs rather than total costs.

May put data onto a more useful scale, ie transform proportions with a logit transform.

Can make comparisons easier by stabilizing variance

Can transform to obtain additivity (ie Cobb-Douglas)

Interaction may only be present due to modeling on the wrong scale, so that transformation erases the need for interaction.

3.2 The power family of transformations

Stretching the axis differentially.

p =	2	1	$\frac{1}{2}$	$\frac{1}{3}$	$-\frac{1}{2}$	-1
	z^2	z	$\sqrt{z}$	$\sqrt[3]{z}$	$\frac{1}{\sqrt{z}}$	$\frac{1}{z}$

Unfortunately does not include ln.

Fix up: consider

$$\frac{z^p - 1}{p}.$$

Take the limit as $p \to 0$ and you get $ln(z)$.

Need to know the shape of these curves.

The most commonly used is probably the log-transform.

Reasons:

Good interpretability in terms of percentage changes.

Turns multiplicative relationships into additive ones.

3.3 Percent change interpretation

Understand interpretations on the log scale, why log transforms result in percentage change interpretations.

Key facts: The log of a product is the sum of the logs.
$ln(1 + \delta) \sim \delta$ for small δ.

Take a log (natural) regression.

$ln(y) = \beta_0 + \beta_1 ln(x)$.

Increase x by δ percent, how does $ln(y)$ shift?

$ln(y^*) = \beta_0 + \beta_1 ln(x \times (1 + \delta))$.

$ln(y^*) = \beta_0 + \beta_1 ln(x) + \beta_1 ln(1 + \delta)$.

$ln(y^*) \sim \beta_0 + \beta_1 ln(x) + \beta_1 \delta$.

How much did $ln(y)$ shift?

$ln(y^*) - ln(y) \sim \beta_1 \delta$.

$ln(y^*/y) \sim \beta_1 \delta$.

$ln(y/y + (y^* - y)/y) \sim \beta_1 \delta$.

$ln(1 + (y^* - y)/y) \sim \beta_1 \delta$.

$(y^* - y)/y \sim \beta_1 \delta$.

Finally: percentage change in y is $\beta_1 \delta$.

4 Multiple regression

4.1 The game plan

1. Model learning curves
2. Model production functions
3. Model costs associated with production function
4. Specialize cost function to include the learning curve model as a special case

4.2 Learning curves

The motivation

Unit costs decrease as cumulative output increases.

Strategic implications for pricing and marketing strategy

Formulation

$$c_t = c_1 n_t^{\alpha_c} e^{u_t},$$

where

- c_t is unit cost in time period t (adjusted for inflation)
- c_1 is unit cost in initial time period
- n_t cumulative production up to but not including time t
- α_c is unit cost elasticity with respect to unit volume
- u_t stochastic disturbance term (our ϵ_t)

Note. Response is unit cost. A multiplicative model.

Make linear by taking logs.

$$\ln(c_t) = \ln(c_1) + \alpha_c \ln(n_t) + u_t.$$

Estimate α_t from a simple regression.

4.3 Cobb Douglas production function.

Model:

$$y = A \times x_1^{\alpha_1} \times x_2^{\alpha_2} \times x_3^{\alpha_3},$$

where

- y is the output
- A denotes the state of technical knowledge
- x_i denotes the quantity of input i
- α_i is the parameter to be estimated (like an elasticity of output with respect to input i)

Note: the response is output. Another multiplicative model.

Define *returns to scale* as $r = \sum \alpha_i$.

4.4 The cost function

The cost function is $C = \sum_i p_i x_i$.

Just the quantity of inputs times their prices.

Relates the minimum cost of producing a level of output y to the prices of the inputs and the state of technical knowledge.

Objective; Find the input levels that minimize the production cost for a given level of output. (Cost minimizer assumption.)

This is an optimization problem, in particular choose input levels to minimize costs. But subject to a **constraint**: the inputs must produce a given level of output, y.

Mathematical technique for solution of constrained optimization: Lagrange multipliers.

It turns out that, assuming the Cobb Douglas production function, then the optimal level of inputs produce a **COST FUNCTION** of the form

$$C = k\, y^{1/r} p_1^{\alpha_1/r} p_2^{\alpha_2/r} p_3^{\alpha_3/r},$$

where

$$k = r(A\alpha_1^{\alpha_1}\alpha_2^{\alpha_2}\alpha_3^{\alpha_3})^{-1/r}.$$

It looks a mess, but notice that it is multiplicative, so taking logs will achieve a linear expression ready for regression.

Further, using the fact that $\alpha_3 = r - \alpha_1 - \alpha_2$ the logged version can be rewritten as

$$\ln(C^*) = \beta_0 + \beta_y \ln(y) + \beta_1 \ln(p_1^*) + \beta_2 \ln(p_2^*),$$

where

- $\ln(C^*) = \ln(C) - \ln(p_3)$
- $\ln(p_1^*) = \ln(p_1) - \ln(p_3)$
- $\ln(p_2^*) = \ln(p_2) - \ln(p_3)$

- $\beta_0 = \ln(k)$
- $\beta_y = 1/r$
- $\beta_1 = \alpha_1/r$
- $\beta_2 = \alpha_2/r$

From this lot we can get at what's of interest, $r, \alpha_1, \alpha_2, \alpha_3$.

4.5 Putting together the Learning Curve and the Cost Function

Objective: make assumptions that incorporate the learning curve into the cost function as a special case.

- Recall that the learning curve equation can be written as

$$\ln(c_t) = \ln(c_1) + \alpha_c \ln(n_t) + u_t.$$

- And the cost equation as

$$\ln(C) = \ln(k)+(1/r)\ln(y)+(\alpha_1/r)\ln(p_1)+(\alpha_2/r)\ln(p_2)+(\alpha_3/r)\ln(p_3)+v_t.$$

Then the question becomes can we put restrictions and assumptions on the cost function so that the learning curve is a special case?

Here's how it goes.

- Define the state of knowledge A_t as $A_t = n_t^{-\alpha_c}$.
- Assume that effects of the input prices are captured by a GNP deflator, ie

$$GNPD_t = (\alpha_1/r)\ln_(p_1) + (\alpha_2/r)\ln_(p_2) + (\alpha_3/r)\ln_(p_3).$$

This leads to a simpler equation:

$$\ln(C_t') = \ln(k') + (\alpha_c/r)\ln(n_t) + 1/r\ln(y_t) + u_t.$$

Here C'_t is a real total cost because it as been adjusted by the GNP deflator.

Finally move to unit real costs rather than total real costs and you obtain

$$\ln(c_t) = \ln(k') + (\alpha_c/r)\ln(n_t) + ((1-r)/r)\ln(y_t) + u_t,$$

which for $r = 1$ is the learning curve model.

How much sense does the previous equation make?

It says that the log of your average real cost at time t depends on two things. 1, how much you have produced up to time t which surrogates for how much knowledge you have and 2, how much you produce at time t as denoted by y_t. If you produce more and your returns to scale are greater than 1 (r 1) then your average cost should decrease – which makes sense.

4.6 Summary

We have seen a variety of econometric models in action.

- There were all **multiplicative**.
- Their functional form was convenient to work with.
- They involved some very strong assumptions.
- Criticism should be tempered by the objective of the modeling.
- They provide a framework and language for discussion rather than a dinner party conversation.