Making predictions

Stephane Hess
stephane.hess@gmail.com

Outline

1. Making predictions: theory
2. Forecasting and correction
3. RP-SP estimation for scale correction

Making predictions: theory

Making predictions: theory

Estimation *vs* prediction

Estimation

- represents step of *understanding* current behaviour
- find best model parameters

Prediction

- how will behaviour change under given scenario?
- e.g. some products are added/removed/changed
- involves applying estimated model, not reestimating it on changed data

Making predictions: theory

The basics of forecasting

- Apply estimated model to:
  - predict choices in **new** settings
  - predict impact of **changes** in products
  - predict impact of **changes** in population
- Mainly relevant in labelled settings

Making predictions: theory

Forecasting matters

- Many studies are primarily interested in willingness-to-pay
- Even then, useful diagnostic check to look at forecasts
- For example, does estimated model give reasonable implied elasticities?
- Well fitting models do not necessarily lead to good forecasts!
- Substantial risk of over-fitting to estimation data

Making predictions: theory

How do we produce forecasts?

- Use estimated β to calculate V_{jn} and P_{jn} for all n and j

	parameter	Apple iPhone	Samsung Galaxy	Huawei P
δ_{Apple}	2.5	1	0	0
$\delta_{Samsung}$	0.75	0	1	0
$\beta_{features}$	0.2	8	4	2
β_{price}	-0.01	600	400	350
	V	-1.9	-2.45	-3.1
	e^V	0.1496	0.0863	0.0450
	P	53.24%	30.72%	16.04%

Question:

What is chosen?

Making predictions: theory

Assigning choice to Apple ignores probabilistic nature

	Apple iPhone	Samsung Galaxy	Huawei P
V	-1.9	-2.45	-3.1
P	53.24%	30.72%	16.04%

- Use average P_j across N instead of individual-level predictions
- Or assign choice according to P_j
  - In our case, take a random draw $0 <= r_U <= 1$
    - if $r_U <= 0.5324$, choose Apple iPhone
    - if $0.5324 < r_U <= 0.8396$, we choose Samsung Galaxy
    - if $0.8396 < r_U$, we choose Huawei P
- Try `deterministic_vs_probabilistic.xlsx` for a mode choice example

Making predictions: theory

Studying the impact of a change in cost

- Carry out prediction with increased cost and get new predicted choices/demand
- Might seem reasonable to compare to base cost choices in data
- Would mean comparing a modelled outcome to an observed outcome
  - model outcomes are affected by error, while data is not
  - model is not likely to perfectly reproduce base scenarios
    - except for linear-in-parameters MNL with full set of ASCs

Making predictions: theory

Solution

- Make two predictions from model
  - Baseline prediction: apply model without changing attributes/population
  - Forecast prediction: apply model with changed data
- Can then compare forecast to baseline application
- Both are affected by the same model bias

Making predictions: theory

Elasticities

- Elasticity is percent change in probability as a result of change in an attribute

Own elasticity of MNL

$$\boxed{E_{i,x_{k,i}} = \frac{\partial V_i}{\partial x_{k,i}} x_{k,i} \left(1 - P_i\left(\beta\right)\right),}$$

with linear in attributes V, $\frac{\partial V_i}{\partial x_{k,i}} = \beta_{x_k}$

Cross-elasticity of MNL

$$\boxed{E_{i,x_{k,j}} = -\frac{\partial V_j}{\partial x_j} x_{k,j} P_j\left(\beta\right),}$$

exhibiting *IIA* characteristic

Much more complex for advanced models

Making predictions: theory

Elasticities: arc elasticity approach

Baseline application

- predicted share for product i: $S_{i,base}$

Forecast prediction

- apply same increase (e.g. cost) for whole sample, say $\Phi_C = \frac{C_{i,new}}{C_{i,base}}$, e.g. $\Phi_C = 1.01$
- obtain new share for product i: $S_{i,forecast}$

Elasticity calculation

- Calculate $E_{i,C} = log\left(\frac{S_{i,forecast}}{S_{i,base}}\right)/log(\Phi_C)$
- See `elasticities.xlsx`

Forecasting and correction

Forecasting and correction

Actual forecasting

- Real world forecasting would likely also require adjusting the sample of respondents
- We often estimate models on small samples, and then apply them to very large samples in sample enumeration
- Another key issue in forecasting is that we also need to forecast changes in the population of decision makers and in the characteristics of the choice sets they face

Forecasting and correction

Sample enumeration

- Assemble a population to use in forecasting
  - either based on real data (e.g., census), or synthetic population
- Apply the model to this data, i.e., make a prediction for each person in that data
- Potentially incorporate weights in that process to make the sample representative
- Then aggregate demand
- If the model incorporates interactions with person characteristics, then the forecasts in sample enumeration will be different depending on those
  - this is one of the key benefits of using deterministic heterogeneity as much as possible

Forecasting and correction

Correcting the scale

- Scale especially in SP data may be very different from real world scale
- Ideally, this can be corrected through joint RP-SP estimation
- But RP data might not always be available
- Can compute implied cost elasticity with estimated coefficients
  - e.g. use a 1% increase in cost and produce forecasts
  - $E = \frac{\log(S_{j,\mathbf{1.01}}/S_{j,b})}{\log(1.01)}$, where $S_{j,b}$ and $S_{j,1.01}$ give market shares before and after the change
- Then adjust the scale so that the cost elasticity is in line with expectations or official guidelines
- See scale_correction.xlsx

Forecasting and correction

Correcting market shares

- Market shares in estimation sample may be very different from real world market shares
- Can recalibrate the model by adjusting the alternative specific constants
- Let α_j^0 be the the estimated constant for alternative j, with S_j^0 being the market share predicted by the model
- Let S_j^1 be the real world market share
- Constants can be adapted to: $\alpha_j^1 = \alpha_j^0 + \ln\left(\frac{S_j^1}{S_j^0}\right)$
- Iterative process which can require a few steps
- See `asc_correction.xlsx`

RP-SP estimation for scale correction

RP-SP estimation for scale correction

Introduction

- Large studies often merge data from different sources
- Examples
  - mode choice and route choice data for the *same* person
  - data from multiple cities or countries
- Larger and possibly more representative samples
- But: potential differences in preferences and scale
  - sensitivities may differ across samples
  - more noise in one dataset than in another
- Not taking this into account may produce biased results

RP-SP estimation for scale correction

Other reason for merging data

- SP is very good at relative valuation
- But:
  - response quality may be an issue
  - sensitivity to attributes may be overstated (or understated)
- Solution: joint RP-SP estimation
  - SP contributes information on relative valuations (useful e.g. for appraisal)
  - RP corrects scale of the model (useful e.g. for elasticities, forecasts)

RP-SP estimation for scale correction

Merging data

- Merge data from different sources into one combined dataset
- Produce utility functions for different alternatives, across all datasets
- In each choice situation, only alternatives from one dataset will be available

RP-SP estimation for scale correction

Sources of differences

- Parameters from all utility functions will be estimated
- Could of course make all parameters dataset-specific
  - would be the same as separate models
  - would not gain anything from joint estimation
- If some parameters are shared across data sources, we can allow for differences in scale (error noise) across sources
- Need to be sure that there are no differences other than just in scale (i.e. MRS is the same)

RP-SP estimation for scale correction

Limiting cases of MNL

- Remember:

$$\boxed{P_i = \frac{e^{\mu V_i}}{\sum_{j=1}^{J} e^{\mu V_j}}}$$

- $\lim_{\mu \to 0} P_i = \frac{1}{J}, \forall i$
- $\lim_{\mu \to \infty} P_i = 1$ if $V_i = \max V_1, ..., V_J$
- But we said μ cannot be separately identified from β

RP-SP estimation for scale correction

With multiple sources, we can allow for scale differences

- Example with two datasets:

$$P_i(D_1) = \frac{e^{\mu_{D_1} V_{i,D_1}}}{\sum_{j=1}^{J_{D_1}} e^{\mu_{D_1} V_{j,D_1}}} \quad \& \quad P_i(D_2) = \frac{e^{\mu_{D_2} V_{i,D_2}}}{\sum_{j=1}^{J_{D_2}} e^{\mu_{D_2} V_{j,D_2}}}$$

- If $V_{\cdot,D_1}$ and $V_{\cdot,D_2}$ share some coefficient(s), possible to estimate either μ_{D_1} or μ_{D_2}
- With D samples, can estimate $D-1$ scale parameters
- Assume we set $\mu_{D_1} = 1$ and estimate μ_{D_2}
  - if $\mu_{D_2} > 1$, modelled choice processes in $D2$ are more deterministic (opposite if $\mu_{D_2} < 1$)
  - scale of β parameters relates to $D1$, relative values of β influenced by both datasets

Questions?

Apollo

www.ApolloChoiceModelling.com

The most flexible choice modelling software (up to a probability)