# **Aggregation and forecasting**

#### Michel Bierlaire

michel.bierlaire@epfl.ch

Transport and Mobility Laboratory





Aggregation and forecasting -p, 1/15



- So far, prediction of individual behavior
- In practice, not useful
- Need for forecast of aggregate demand:
  - number of trips
  - number of passengers
  - etc.







Linear models

$$t_n = \alpha + \beta p_n$$

where

- $t_n$ : number of trips from zone n
- $p_n$ : population in zone n
- If  $\bar{p}$  is the average population
- $\bar{t} = \alpha + \beta \bar{p}$  is the average number of trips
- It does not work with choice models, because they are nonlinear







• "Travel/no travel" model,  $y_n$  income

No travel 
$$V_1 = 0$$
  
Travel  $V_2 = -3 + 3y_n$ 

	Income	V1	V2	P1	P2
Household 1	1	0	0	50%	50%
Household 2	10	0	27	0%	100%
Avg. income	5.5	0	13.5	0%	100%
Avg. probabilities				25%	75%





Addregation and forecasting -p, 4/15



• Choice model

$$P(i|x_n)$$

where  $x_n$  gathers attributes of all alternatives and socio-economic characteristics of n

• If the population is composed of N individuals, the total expected number of individuals choosing i is

$$N(i) = \sum_{n=1}^{N} P(i|x_n)$$

- Hopeless to know  $x_n$  for every and each individual
- The sum would involve a lot of terms.
- The distribution of *x* could be used.







- Assume that the distribution of x is continuous with PDF p(x)
- Then the share of the population choosing i is given by

$$\widehat{W}(i) = \int_{x} P(i|x)p(x)dx$$

- In practice, p(x) is also unknown
- The integral may be cumbersome to compute







- If the population is segmented in *S* homogeneous segments
- If  $N_s$  is the number of individuals in segment s
- Then

$$\widehat{N}(i) = \sum_{s=1}^{S} N_s P(i|x_s)$$





## Illustration

The travel model:

• "Travel/no travel" model,  $y_n$  income

$$P(\text{travel}) = \frac{e^{-3+3y_n}}{1+e^{-3+3y_n}}$$

- Population: N = 200'000 persons
- Sample: S = 500 persons
- Sampling rate: S/N = 1/400





## Illustration

S	$y_s$	$S_s$	$N_s$	P(travel)	$PS_s$	$PN_s$
1	0	150	20000	4.7%	7	949
2	0.5	200	30000	18.2%	36	5473
3	1	40	50000	50.0%	20	25000
4	1.5	10	50000	81.8%	8	40879
5	2	50	30000	95.3%	48	28577
6	2.5	50	20000	98.9%	49	19780
		500	200000		169	120657
	120657  eq 400  imes 169 = 67542					

People with low probability of travel are oversampled





Aggregation and forecasting -p, 9/15



Most practical method: sample enumeration

- Let n be an observation in the sample belonging to segment s
- Let  $W_s$  be the weight of segment s, that is

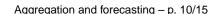
 $W_s = \frac{N_s}{S_s} = \frac{\text{\# persons in segment } s \text{ in population}}{\text{\# persons in segment } s \text{ in sample}}$ 

• The number of persons choosing alt. *i* is estimated by

$$\widehat{N}(i) = \sum_{n \in \text{sample}} \sum_{s} W_s P(i|x_n) I_{ns}$$

where  $I_{ns} = 1$  if individual *n* belongs to segment *s*, 0 otherwise







### We can write

$$\widehat{N}(i) = \sum_{\substack{n \in \text{sample} \\ n \in \text{sample}}} \sum_{s} W_s P(i|x_n) I_{ns}$$
$$= \sum_{\substack{n \in \text{sample} \\ s}} P(i|x_n) \sum_{s} W_s I_{ns}$$

The term  $\sum_{s} W_{s}I_{ns}$  is the weight of individuals *n* belonging to segment *s*.

The share of alt. *i* is estimated by W(i) =

$$\frac{1}{N}\sum_{n\in\text{sample}} P(i|x_n)\sum_s W_s I_{ns} = \sum_{n\in\text{sample}} P(i|x_n)\sum_s \frac{N_s}{N}\frac{1}{S_s}I_{ns}$$





## Forecasting

- Modify  $x_n$  in the sample to reflect anticipated modifications
- Apply the sample enumeration again





## Example

S	$y_s$	$S_s$	P(travel)	$W_s$	Trips
1	0	150	4.74%	133.33	949
2	0.5	200	18.24%	150	5473
3	1	40	50.00%	1250	25000
4	1.5	10	81.76%	5000	40879
5	2	50	95.26%	600	28577
6	2.5	50	98.90%	400	19780
					120657

- Increase all salaries by 0.5
- What is the impact on the total number of trips?





## Example

S	$y_s$	$S_s$	P(travel)	$W_s$	Trips
1	0.5	150	18.24%	133.33	3649
2	1	200	50.00%	150	15000
3	1.5	40	81.76%	1250	40879
4	2	10	95.26%	5000	47629
5	2.5	50	98.90%	600	29670
6	3	50	99.75%	400	19951
					156777

- Before: 120657
- After: 156777
- Increase: about 30%





## **Summary**

- Aggregation:
  - Sample enumeration.
  - Correct for sampling errors using weights.
- Forecasting:
  - Forecast the value of the explanatory variables x.
  - Aggregate.



