Multinomial Logit and WTP

Gianfranco Atzeni

Environmental Economics

Families of probabilistic DCM

RUT leads to families of probabilistic discrete choice models that describe how choice probabilities respond to changes in choice options (or equivalently, their attributes) and/or covariates representing differences in individuals.

The probability that individual i chooses option j from a set of competing options is:

$$P(y_i = j) = Pr(U_{ij} \geq U_{ik}) \ \ for\ all\ k$$

for all j options in choice set C_j

...Families of probabilistic DCM...

$$
\begin{aligned}
P(y_i = j) &= Pr(U_{ij} \geq U_{ik}) \ \textit{for all } k \\
&= Pr(V_{ij} + \varepsilon_{ij} \geq V_{ik} + \varepsilon_{ik}) \ \textit{for all } k \\
&= Pr(\varepsilon_{ik} - \varepsilon_{ij} \leq V_{ij} - V_{ik}) \ \textit{for all } k
\end{aligned} \tag{1}
$$

Equation (1) says that the probability that individual i chooses option $y = j$ from a given choice set equals the probability that the systematic and random components of option j for individual i are larger than the systematic and random components of all other options competing with option j

...Families of probabilistic DCM

Different probabilistic discrete choice models can be derived from equation (1) by making different assumptions about probability distributions for ε_{in}, as for example:

distributed as non-independent, non-identically distributed normal random variates (Thurstone);

Independent Identically Distributed (IID) Gumbel (McFadden). The Gumbel distribution closely resembles the normal but is slightly asymmetric;

it has the advantage of yielding closed form expressions for the choice probabilities if random components are IID, i.e. multinomial logit (MNL) model (also known as a conditional logit model) used in practical applications.

Multinomial Logit

The outcome, y_i, for individual i is one of the m alternatives. As indicated above, we set y_i if the outcome is the jth alternative, with $j = 1, 2, ..., m$.

The values $1, 2, ..., m$ are arbitrary and the same regression results are obtained if we use values $3, 5, 8, ...$

The ordering of the values also does not matter, unless we use an ordered model.

The probability that the outcome for individual i is alternative j, conditional on the regressors x_i, is

$$p_{ij} = Pr(y_i = j) = F_j(x_i, \theta), \quad j = 1, ..., m, \quad i = 1, ..., N \tag{2}$$

$F_j(.)$ correspond to different multinomial models.

MNL - Normalization

Only $m-1$ of the probabilities can be freely specified because probabilities sum to one.

For example $F_m(x_i,\theta) = 1 - \sum_{j=1}^{m-1} F_j(x_i,\theta)$

Therefore, MNL models require normalization.

MNL - Marginal Effects

The parameters of MNL models are generally not directly interpretable. A positive coefficient need not mean that an increase in the regressor leads to an increase in the probability of an outcome being selected.

We need to compute marginal effects.

For individual i the ME of a change in the kth regressor on the probability that the alternative j is the outcome is:

$$ME_{ijk} = \frac{\partial Pr(y_i = j)}{\partial x_{ik}} = \frac{\partial F_j(x_i, \theta)}{\partial x_{ik}} \tag{3}$$

For each regressor there will be m MEs corresponding to the m probabilities.

As for other non linear models, these marginal effects vary with the evaluation point x.

Case specific and alternative-specific

case-specific or alternative-invariant regressors: are regressors, such as gender, individual income etc., that do not vary across alternatives;

alternative-specific or case-varying regressors: are regressors, such as price, that may vary across alternatives.

Example

Data on individual choice of whether to fish using for possible modes:

from the beach;

from the pier;

from a private boat;

from a charter boat. One explanatory variable is case specific (`income`) and the others (`price` and `crate`, catch rate) are alternative specific.

Data description

```
Contains data from mus15data.dta
  obs:         1,182
 vars:            16                      26 Nov 2008 17:16
 size:        75,648
------------------------------------------------------------------------
              storage  display    value
variable name   type   format     label      variable label
------------------------------------------------------------------------
mode            float  %9.0g      modetype   Fishing mode
price           float  %9.0g                 price for chosen alternative
crate           float  %9.0g                 catch rate for chosen alternative
dbeach          float  %9.0g                 1 if beach mode chosen
dpier           float  %9.0g                 1 if pier mode chosen
dprivate        float  %9.0g                 1 if private boat mode chosen
dcharter        float  %9.0g                 1 if charter boat mode chosen
pbeach          float  %9.0g                 price for beach mode
ppier           float  %9.0g                 price for pier mode
pprivate        float  %9.0g                 price for private boat mode
pcharter        float  %9.0g                 price for charter boat mode
qbeach          float  %9.0g                 catch rate for beach mode
qpier           float  %9.0g                 catch rate for pier mode
qprivate        float  %9.0g                 catch rate for private boat mode
qcharter        float  %9.0g                 catch rate for charter boat mode
income          float  %9.0g                 monthly income in thousands $
------------------------------------------------------------------------
```

Summary statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
mode	1182	3.005076	.9936162	1	4
price	1182	52.08197	53.82997	1.29	666.11
crate	1182	.3893684	.5605964	.0002	2.3101
dbeach	1182	.1133672	.3171753	0	1
dpier	1182	.1505922	.3578023	0	1
dprivate	1182	.3536379	.4783008	0	1
dcharter	1182	.3824027	.4861799	0	1
pbeach	1182	103.422	103.641	1.29	843.186
ppier	1182	103.422	103.641	1.29	843.186
pprivate	1182	55.25657	62.71344	2.29	666.11
pcharter	1182	84.37924	63.54465	27.29	691.11
qbeach	1182	.2410113	.1907524	.0678	.5333
qpier	1182	.1622237	.1603898	.0014	.4522
qprivate	1182	.1712146	.2097885	.0002	.7369
qcharter	1182	.6293679	.7061142	.0021	2.3101
income	1182	4.099337	2.461964	.4166667	12.5

Note that the variable `mode` takes on the values from 1 to 4 depending on the fishing mode.

Summary statistics

Fishing mode	Freq.	Percent	Cum.
beach	134	11.34	11.34
pier	178	15.06	26.40
private	418	35.36	61.76
charter	452	38.24	100.00
Total	1,182	100.00	

```
. label list
modetype:
           1 beach
           2 pier
           3 private
           4 charter
```

Summary statistics

For the case-specific regressor `income` this summarize the relationship between it and the dependent variable.

```
. table mode, contents(N income mean income sd income)
```

Fishing mode	N(income)	mean(income)	sd(income)
beach	134	4.051617	2.50542
pier	178	3.387172	2.340324
private	418	4.654107	2.777898
charter	452	3.880899	2.050029

Summary statistics

For the alternative-specific regressor `price` this summarize the relationship between it and the dependent variable.

```
. table mode, contents(mean pbeach mean ppier mean pprivate mean pcharter)
```

Fishing mode	mean(pbeach)	mean(ppier)	mean(pprivate)	mean(pcharter)
beach	35.69949	35.69949	97.80914	125.0032
pier	30.57133	30.57133	82.42908	109.7634
private	137.5271	137.5271	41.60681	70.58408
charter	120.6483	120.6483	44.56376	75.09694

Summary statistics

For the alternative-specific regressor `crate` (catch rate) this summarize the relationship between it and the dependent variable.

```
. table mode, contents(mean qbeach mean qpier mean qprivate mean qcharter)
```

Fishing mode	mean(qbeach)	mean(qpier)	mean(qprivate)	mean(qcharter)
beach	.2791948	.2190015	.1593985	.5176089
pier	.2614444	.2025348	.1501489	.4980798
private	.2082868	.1297646	.1775412	.6539167
charter	.2519077	.1595341	.1771628	.6914998

MLN

The Multinomial Logit model can be used when all the regressors are case specific. The deterministic part of the random utility is modeled as:

$$V_{ij} = x_i' \beta_j \quad i = 1, ..., n \;\; and \quad j = 1, ..., m. \tag{4}$$

where x_i is a m-dimensional vector of characteristics of individual i and β_j is an m-dimensional parameter vector specific to alternative j.

Assuming that ε has an independent type-I extreme value distribution, the model is:

$$p_{ij} = \frac{e^{x_i' \beta_j}}{\sum_{l=1}^{m} e^{x_i' \beta_l}} \quad j = 1, ..., m. \tag{5}$$

where x_i are case specific regressors (in our example `income` and the intercept).

β_j is set to zero for one of the categories, and coefficients are interpreted with respect to that category (**base category**).

MLN estimation

We regress fishing mode on an intercept and income, the only case-specific regressor. The `beach` category is set as base category.

```
Multinomial logistic regression                 Number of obs     =       1182
                                                LR chi2(3)        =      41.14
                                                Prob > chi2       =     0.0000
Log likelihood = -1477.1506                     Pseudo R2         =     0.0137
```

mode	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
beach	(base outcome)					
pier						
income	-.1434029	.0532884	-2.69	0.007	-.2478463	-.0389595
_cons	.8141503	.228632	3.56	0.000	.3660399	1.262261
private						
income	.0919064	.0406637	2.26	0.024	.0122069	.1716058
_cons	.7389208	.1967309	3.76	0.000	.3533352	1.124506
charter						
income	-.0316399	.0418463	-0.76	0.450	-.1136571	.0503774
_cons	1.341291	.1945167	6.90	0.000	.9600457	1.722537

Test of significance

Two of the three coefficient estimates of `income` are statistically significant (5% level). However, it is important to notice that the results of individual testing vary with the omitted category.

We can perform a joint test, using a Wald test:

```
. test income

 ( 1)  [beach]o.income = 0
 ( 2)  [pier]income = 0
 ( 3)  [private]income = 0
 ( 4)  [charter]income = 0
       Constraint 1 dropped

           chi2(  3) =   37.70
         Prob > chi2 =    0.0000
```

Income is highly significant

Coefficient intepretation

Coefficients in a MNL model can be interpreted in the same way as binary logit models parameters are interpreted, with respect to base category.

MNL is equivalent to a series of pairwise logit models. If the base category is the first category, then MLN implies that:

$$Pr(y_i = j | y_i = j \quad or \quad 1) = \frac{Pr(y_i = j)}{Pr(y_i = j) + Pr(y_i = 1)} = \frac{e^{(x_i' \beta_j)}}{1 + e^{(x_i' \beta_j)}} \tag{6}$$

$\hat{\beta}_j$ positive means that as the regressor increases we are more likely to choose alternative j than alternative 1. This interpretation vary with the base category.

Predicted probabilities

We can compute the average predicted probability of a given outcome (for example outcome 3 - private boat fishing)

```
. margins, predict(outcome(3)) noatlegend

Predictive margins                          Number of obs    =       1182
Model VCE    : OIM

Expression   : Pr(mode==private), predict(outcome(3))
```

	Margin	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
_cons	.3536379	.0137114	25.79	0.000	.326764	.3805118

Marginal effects

For an unordered MNL model there is no single conditional mean of the dependent variable y, instead there are m alternatives. We want to know how these probabilities change as regressors change.

The MEs are:

$$p_{ij} = \frac{e^{x_i'\beta_j}}{\sum_{l=1}^{m} e^{x_i'\beta_l}}$$
$$\frac{\partial p_{ij}}{\partial x_i} = p_{ij}(\beta_j - \sum_l p_{il}\beta_l) \tag{7}$$

The MEs vary with the point of evaluation x_i, because p_{ij} varies with x_i.

The sign of the regression coefficients do not give the signs of the MEs. For the variable x the ME is positive if $\beta_j > \sum_l p_{il}\beta_l$

Marginal effects

```
. predict p1 p2 p3 p4, pr

.
. gen me3=p3*(_b[3:income]-(p2*_b[2:income]+p3*_b[3:income]+p4*_b[4:income]))

. sum me3
```

Variable	Obs	Mean	Std. Dev.	Min	Max
me3	1182	.0317562	.0015078	.0279562	.0336326

One unit change in `income` increases by 0.032 the probability of fishing from a private boat rather than from a beach, pier, or charter boat.

Marginal effects

Alternatively using the command `margins`:

```
. margins, dydx(*) predict(outcome(3)) atmean noatlegend

Conditional marginal effects                    Number of obs   =       1182
Model VCE    : OIM

Expression   : Pr(mode==private), predict(outcome(3))
dy/dx w.r.t. : income
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
income	.0325985	.005692	5.73	0.000	.0214424	.0437547

One unit change in `income` increases by 0.032 the probability of fishing from a private boat rather than from a beach, pier, or charter boat.

Conditional logit

Some multinomial studies use datasets that include also alternative-specific variables, such as in our example prices or catching rate for all the alternatives, and not just for the chosen alternative.

Data need to be in long form, with one observation providing the data for just one alternative for an individual. The dataset in our example it is not in this format, with one observation containing data for all four alternatives for an individual.

Long-form from wide-form data

For example for the first observation of the dataset:

```
. list mode price pbeach ppier pprivate pcharter in 1, clean

          mode    price   pbeach    ppier   pprivate   pcharter
  1.   charter   182.93   157.93   157.93     157.93     182.93
```

It has data for the price of all four observations. The chosen mode was `charter`, so `price` was set to equal `pcharter`. Using the `reshape` command, we transform into a form with four observation for each individual according to whether the suffix is `beach`, `pier`, `private`, or `charter`

```
. reshape long d p q, i(id) j(fishmode beach pier private charter) string

Data                               wide   ->   long
-----------------------------------------------------------------------------
Number of obs.                     1182   ->    4728
Number of variables                  26   ->      18
j variable (4 values)                     ->   fishmode
xij variables:
              dbeach dpier ... dcharter   ->   d
              pbeach ppier ... pcharter   ->   p
              qbeach qpier ... qcharter   ->   q
-----------------------------------------------------------------------------
```

Alternative-specific conditional logit

When some regressors are alternative-specific we use the **conditional logit** model (also termed as **mixed logit**):

$$p_{ij} = \frac{e^{(x'_{ij}\beta + z'_i\gamma_j)}}{\sum_{l=1}^{m} e^{(x'_{il}\beta + z'_i\gamma_l)}} \quad j = 1, ..., m. \tag{8}$$

where x_{ij} are alternative-specific regressors and z_i are case-specific regressors. For model identification one of the γ_j is set to zero as for the MNL model.

asclogit - Estimation

```
. asclogit d p q, case (id) alternatives(fishmode) casevars(income) basealternative(beach) nolog

Alternative-specific conditional logit         Number of obs      =      4728
Case variable: id                              Number of cases    =      1182

Alternative variable: fishmode                 Alts per case: min =         4
                                                              avg =       4.0
                                                              max =         4

                                                  Wald chi2(5)    =    252.98
Log likelihood = -1215.1376                       Prob > chi2     =    0.0000
```

d	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
fishmode						
p	-.0251166	.0017317	-14.50	0.000	-.0285106	-.0217225
q	.357782	.1097733	3.26	0.001	.1426302	.5729337
beach	(base alternative)					
charter						
income	-.0332917	.0503409	-0.66	0.508	-.131958	.0653745
_cons	1.694366	.2240506	7.56	0.000	1.255235	2.133497
pier						
income	-.1275771	.0506395	-2.52	0.012	-.2268288	-.0283255
_cons	.7779593	.2204939	3.53	0.000	.3457992	1.210119
private						
income	.0894398	.0500671	1.79	0.074	-.0086898	.1875694
_cons	.5272788	.2227927	2.37	0.018	.0906132	.9639444

Coefficient interpretation

We denote the alternative-specific regressors by x_r with the coefficient β_r. The effect of a change in x_{rik}, which is the value of x_r for individual i and alternative k is

$$\frac{\partial p_{ij}}{\partial x_{ri}} = \begin{cases} p_{ij}(1 - p_{ij})\beta_r & j = k \\ -p_{ij}p_{ik}\beta_r & j \neq k \end{cases} \quad (9)$$

If $\beta_r > 0$ then the own effect is positive because $p_{ij}(1 - pij)\beta_r > 0$, and the cross-effect is negative because $-p_{ij}p_{ik}\beta_r < 0$. So a positive coefficient means that if the regressor increases for one category, then the category is chosen more and the other categories are chosen less; viceversa for negative coefficient.

Coefficient interpretation

In our example the negative price coeffecent of -0.025 means that if the price of one mode of fishing increases, then the demand for that mode decreases and the demand for all other modes increases.

For catch rate (q) the positive coefficient of 0.36 means that a higher catch rate for one mode of fishing increases the demand for that mode and decreases the demand for the other modes.

Coefficients of case-specific regressors are interpreted as parameter of a binary logit. The income coefficients of -0.033, -0.128, 0.089 mean that relative to the probability of beach fishing an increase in income leads to a decrease in the probability of charter boat and pier fishing, and an increase in the probability of private boat fishing.

MEs

We compute the ME for the regressor `price`

```
. estat mfx, varlist(p)
```

Pr(choice = beach|1 selected) = .05248806

variable	dp/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
p							
beach	-.001249	.000121	-10.29	0.000	-.001487	-.001011	103.42
charter	.000609	.000061	9.97	0.000	.000489	.000729	84.379
pier	.000087	.000016	5.42	0.000	.000055	.000118	103.42
private	.000553	.000056	9.88	0.000	.000443	.000663	55.257

Pr(choice = charter|1 selected) = .46206853

variable	dp/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
p							
beach	.000609	.000061	9.97	0.000	.000489	.000729	103.42
charter	-.006243	.000441	-14.15	0.000	-.007108	-.005378	84.379
pier	.000764	.000071	10.69	0.000	.000624	.000904	103.42
private	.00487	.000452	10.77	0.000	.003983	.005756	55.257

MEs

Pr(choice = pier|1 selected) = .06584968

variable	dp/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
p							
beach	.000087	.000016	5.42	0.000	.000055	.000118	103.42
charter	.000764	.000071	10.69	0.000	.000624	.000904	84.379
pier	-.001545	.000138	-11.16	0.000	-.001816	-.001274	103.42
private	.000694	.000066	10.58	0.000	.000565	.000822	55.257

Pr(choice = private|1 selected) = .41959373

variable	dp/dx	Std. Err.	z	P>|z|	[95% C.I.	]	X
p							
beach	.000553	.000056	9.88	0.000	.000443	.000663	103.42
charter	.00487	.000452	10.77	0.000	.003983	.005756	84.379
pier	.000694	.000066	10.58	0.000	.000565	.000822	103.42
private	-.006117	.000444	-13.77	0.000	-.006987	-.005246	55.257

MEs

The header of the first section of the `estate mfx` output gives:

$p_{11} = Pr(choice = beach | one\ \ choice\ \ is\ \ selected) = 0.0525$. Using 9 and the estimated coefficient -0.0251 we can estimate the own-effect as $0.0525 \times (1 - 0.0525) \times (-0.0251) = -0.001249$, which is the first ME given in the output.

This means that a \$1 increase in the price of beach fishing decreases the probability of beach fishing by 0.001249 for a fictional observation with `p`, `q`, and `income` set to sample mean.

The second value 0.000609 means that \$1 increase in the price of charter boat fishing increases beach fishing by 0.000609. And so on.

Random-parameters logit

The random-parameters (RP) or mixed logit model (MLM) relaxes the IIA assumption (Independence of Irrelevant Alternatives).

IIA assumption states that: If A is preferred to B out of the choice set {A,B}, introducing a third option X, expanding the choice set to {A,B,X}, must not make B preferable to A.

MLM allows the parameters in the Conditional logit (CL) to be normally distributed.

MLM

The MLM or RP model is obtained as above from the random utility model assuming that the error ε_{ij} are type II extreme-value distributed, like in the CL model, and the parameters β and γ_i are normally distributed.

The utility of alternative j is:

$$\begin{aligned} U_{ij} &= x'_{ij}\beta_i + z'_i\gamma_{ji} + \varepsilon_{ij} \\ &= x'_{ij}\beta + z'_i\gamma_j + x'_{ij}v_i + z'_i w_{ji} + \varepsilon_{ij} \end{aligned} \tag{10}$$

where x_{ij} are the alternative-specific regressors, z_i are the case specific regressors, $\beta_i = \beta + v_i$ and $v_i \sim N(0, \Sigma_\beta)$ and $\gamma_{ji} = \gamma_j + w_{ji}$, and $w_{ji} \sim N(0, \Sigma_{\gamma_j})$.

We have now a combined error term $(x'_{ij}v_i + z'_i w_{ji} + \varepsilon_{ij})$ which is correlated across alternatives, while the ε_{ij} alone were not.

MLM

The CL model is then:

$$p_{ij}|v_i, w_{ij} = \frac{e^{x'_{ij}\beta_j + z'_i\gamma_j + x'_{ij}\beta + z'_i w_{ji}}}{\sum_{l=1}^{m} e^{x'_{ij}\beta_j + z'_i\gamma_j + x'_{ij}v_i + z'_i w_{ji}}} \tag{11}$$

The Maximum Likelihood estimator is based on p_{ij}, which requires integrating out v_i and w_{ij}.

STATA estimates RPM or MLM using a frequency simulator that makes several draws of v_i and w_{ij} from the normal distribution.

This is implemented in STATA by the command `mixlogit`, which however has no option for case specific regressors unlike `asclogit`. Therefore, we need to manually create the variables for the case specific regressors and intercept.

Data preparation

To make data ready for the `mixlogit` command we need:

a normalization: we set $\gamma_{pier} = 0$;

to construct 3 intercepts and 3 interactions between fishing mode and `income`

```
. gen dbeach=fishmode=="beach"

. gen dprivate=fishmode=="private"

. gen dcharter=fishmode=="charter"

. gen ybeach=dbeach*income

. gen yprivate=dprivate*income

. gen ycharter=dcharter*income
```

Estimation with `mixlogit`

The parameter for `p` are specified to be random (`rand` option) all other are fixed (`q beach private ybeach yprivate`)

```
. mixlogit d q dbeach dprivate ybeach yprivate, group(id) rand(p)

Iteration 0:   log likelihood = -602.33588  (not concave)
Iteration 1:   log likelihood = -447.46151
Iteration 2:   log likelihood = -435.30285
Iteration 3:   log likelihood = -434.56116
Iteration 4:   log likelihood = -434.52856
Iteration 5:   log likelihood = -434.52844
Iteration 6:   log likelihood = -434.52844

Mixed logit model                               Number of obs   =       2190
                                                LR chi2(1)      =      64.57
Log likelihood = -434.52844                     Prob > chi2     =     0.0000
```

d	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Mean						
q	.7840088	.9147869	0.86	0.391	-1.008941	2.576958
dbeach	-.7742955	.224233	-3.45	0.001	-1.213784	-.3348069
dprivate	-.212556	.3059978	-0.69	0.487	-.8123006	.3871886
ybeach	.1199613	.0492249	2.44	0.015	.0234822	.2164404
yprivate	.1717711	.0716575	2.40	0.017	.0313251	.3122172
p	-.1069866	.0274475	-3.90	0.000	-.1607827	-.0531904
SD						
p	.0598364	.0191597	3.12	0.002	.022284	.0973888

The sign of the estimated standard deviations is irrelevant: interpret them as

Note that the option `group()` is used to identify each case or individual, paralleling the option `case()` of the `asclogit` command.

As far as the interpretation of the random parameters for `p` we note that: there is a considerable variation across individuals in the effect of price. The random coefficients have a mean of -0.107 and a standard deviation of 0.059, both statistically significant at 5% level.

WTP / WTA...

- Whatever Random Utility Model has been chosen is estimated by econometric analysis to provide the mean and median WTP/WTA information, via the estimation of indirect utility function.
- Welfare measurement:
  - assess how respondents' utility would change if the attributes of a non-marketed good were changed from their current level (the status quo) to some different level (the levels set by the policy option under review);
  - express this utility change in money terms.

...WTP / WTA

It is possibile to calculate marginal WTP for any one of the option attributes according to:

$$MWTP = -\frac{b_k}{\beta} \qquad (12)$$

- where β is the coefficient on cost and b_k is the coefficient on attribute k.
- **Interpretation**: b_k is the utility from an extra unit of attribute k; β is the money value of an extra unit of utility.
- Their ratio: the monetary value of the utility coming from an extra unit of attribute k (**IMPLICIT PRICE** of the attribute).

Validity

The central problem in assessing the validity of WTP/WTA values obtained from any stated preference study is the absence of a definitive yardstick against which to compare those measures.

- Not necessary the case in all survey research (for example election opinion polls).
- It is generally a problem for non-market goods since, with very few exceptions, actual values are unobservable.

Residents and Tourist trade-offs

Nanni Concu, Gianfranco Atzeni (2012) **Conflicting preferences among tourists and residents**, Tourism Management.

The paper studies the complex relationship between tourism and host environments

- Tourism development depends on features and quality of natural, cultural, and heritage resources, among other things.
- Tourism also demands services and goods that could alter these environments, and hence it has the potential to degrade the resources on which its development is based.
- Comparing the benefits received and the costs incurred by host communities and tourists is then necessary to determine the optimal level of tourist development.

Benefits and costs

Benefits and costs of tourism development are difficult to quantify because of positive and negative externalities, non-market values, and opportunity costs.

- For instance, tourists incur costs by both purchasing the holiday package and suffering from overcrowding of tourist facilities.
- The host community could gain from tourism revenues and from revitalising of local traditions, and incur costs such as disruption of social relations and environmental degradation

Aim of the study

The aim is to provide a monetary estimation of residents' and tourists' perceptions of the impacts of recreational uses of resources in host communities

We use a stated preference technique, the Discrete Choice Experiments

- Study area: Alghero
- Coastal development policies have been the source of serious political debate in recent years
- Regulation: Sustainable Tourism Development Regional Plan (Piano Regionale di Sviluppo Turistico Sostenibile, PRSTS), and the Regional Landscape Plan (Piano Paesaggistico Regionale, PPR).

Regulation...

- The PRSTS represents the first strategic plan for the preservation and use of natural resources for tourism purposes in Sardinia
- Coastal development policies have been the source of serious political debate in recent years
- The PPR is the main instrument for implementing conservation and protection measures. Two principles guided the 2004 framework:
    - sustainable development;
    - homogeneity of planning processes.
- The major prescription of the PPR is the total ban of new buildings and infrastructures within 2 km from the seashore.

On one hand, this regulation aimed at protecting the coastal landscape and environment; on the other hand, it posed extensive limits on the use of this resource for tourism development.

...Regulation

Another important regulation of the 2004 reform includes a set of incentives for the renovation and restoration of old suburbs in tourist towns. This measure aimed at reducing the spread of tourist infrastructures, and making better use of the existing urban assets

- An unintended consequence of such a measure is the concentration of tourists in urban areas, with increased **congestion** affecting both residents' quality of life and the recreational experience.
- Alghero is an ideal place to investigate on communities and tourist preferences, and the potential conflicts between them.

Survey design and administration...

The choice modelling application was organized as follows.

1. preliminary questionnaire to interview a random sample of residents
   - open and Likert-type questions on the effects of tourism on natural resources, public services, resident's quality of life, and on other socio-economic indicators.
2. Out of this information, we defined four attributes to describe the main perceived effects of tourism and to create various scenarios of tourism development.
3. The preliminary questionnaire contained some questions on policies for promotion and management of tourist flows, and on how to fund these policies.
4. willingness to pay for the proposed policies. This aimed to provide a range of monetary values for the final questionnaire.

...Survey design and administration

We used this information to design the choice modelling questionnaire. This questionnaire has three parts.

1. contains a set of questions on respondents' attitudes, and some reminders of income constraints on individual and public choices.
2. contains the choice sets generated by combining five levels of the four attributes.
3. These combinations describe the claimed effects of PPR (the **status quo** at the time) and hypothetical alternatives.

Status quo

The status quo levels describe the expected effects of the PPR, based on the policy makers' claim and past trends of the tourism industry:

- there was still a great deal of uncertainty regarding the true impact of the PPR;
- the time frame proposed to residents was five years, that is, the proposed changes take place every year for five years.

Attributes and levels

Table 3
Choice attributes.

Attributes	Description	Label	Levels
Distance of new buildings from the seashore	Level of protection assigned to the coastal environment	DIST	• 150 m • 500 m • 1000 m • 2000 m (Status quo) • 3000 m
Number of new jobs in the tourist sector (per year in Alghero)	Economic impact of tourism	EMPL	• 0 • 20 • 40 (Status quo) • 60 • 80
Increase of time required by daily activities (in minutes)	Impact on residents' quality of life of increased tourist flows	TIME	• 5 • 10 • 15 (Status Quo) • 30 • over 30
Payment vehicle	Residents: • local tax increase (absolute value in euros per year)	COST	• 0 (Status quo) • 10 • 20 • 30 • 40
	Tourists • increase of cost of holidaying in Alghero (absolute value in euros per day)		• 0 (Status quo) • 2 • 4 • 6 • 8

Results

Table 5
Random parameter models with mean heterogeneity.

Variables	Residents		Tourists	
	Coeff.	Standard error	Coeff.	Standard error
Non-random parameters (b)				
Alternative specific constant	−0.6477***	0.101	−0.9290***	0.246
Random parameters of choice attributes (b_{ij})				
Distance of new building from the seashore (DIST)	5.17E−05	0.788E−04	3.67E−04***	0.000
Number of new jobs in the tourist sector (EMPL)	0.0796***	0.009	0.0910***	0.019
Increase of congestion (TIME)	−0.0452***	0.009	−0.0331	0.037
Local tax increase (in euro) (COST)	−0.0276***	0.006		
Increase of cost of the holiday (in euro) (COST)			−0.063	0.064

Results

Heterogeneity in mean (d_{ij})				
DIST:gender (female = 1)	0.0005***	0.000		
DIST:income from tourism (yes = 1)	−0.0004***	0.000		
DIST:# of dependants			−0.0002 ***	0.779976E−04
EMPL:age	−0.0006***	0.000		
EMPL:income			−6.01E−7***	0.283362E−06
TIME:income from tourism (yes = 1)	−0.0324***	0.014		
TIME:age			−0.0024***	0.001
COST:gender (female = 1)	−0.0096***	0.005		
COST:income from tourism (yes = 1)	0.0155***	0.002		
COST:# of dependants	−0.0048***	0.006		
COST:age			−0.0047***	0.002
Derived standard deviations of parameter distributions (w_{iz})				
NsDIST	0.0015***	0.000	0.0003	0.000
NsEMPL	0.0146	0.010	0.0791***	0.023
NsTIME	0.0639***	0.024	0.1299***	0.050
NsCOST	0.0450***	0.013	0.14	0.104
Pseudo *Rsq*	0.237		0.338	
Log likelihood	−3306.55		−997.75	
Observations	6264		2188	
% Correct predictions	72.9		77.7	

***Significant at 1%, **significant at 5%; *significant at 10%.

Implicit prices

Table 6
Implicit prices.

Attribute	Residents (income from tourism = no)	
	IP (€)	95% CI
Distance of new building from the seashore	0.013***	0.007–0.019
Number of new jobs in the tourist sector	1.95***	1.374–2.527
Increase of congestion	−1.11***	−1.576 to −0.642

***Significant at 1%, **significant at 5%; *significant at 10%.

Implicit prices

Residents (income from tourism = yes)		Tourists	
IP (€)	95% CI	IP (€)	95% CI
0.003	−0.009 to 0.015	0.005	−0.004 to 0.015
3.15***	1.539–4.753	1.28	−0.997 to 3.552
−3.07***	−4.838 to −1.303	−0.46	−1.631 to 0.702

Conflicting preferences

Results indicate that there are several conflicting preferences over tourism development:

- increasing the level of environmental conservation in Alghero provides welfare gains only to those residents that do not earn an income from tourism;
- both tourists and tourist service providers are not concerned about environmental protection.
- congestion seems to negatively affect Alghero residents e including those that earn their income from tourism.
- tourists do not indicate they are impacted by congestion.

jobs

- Increasing employment opportunities is very important for residents, while tourists have no concern for the impact of tourist development on local jobs.
- Our study shows that tourists are not sensitive to changes in the cost of their holidays in the price range used in the analysis (from 0 to 8 euros per day). Hence the local authorities could rise revenues at the cost of only a small reduction of the number of tourists and their length of stay.