Economic Valuation of Environmental Change

Module 5.1: Choice Experiments: Modeling

Book chapters: PR Ch. 16, CBB CH. 5

LaTex commands

In [1]:
```
# Activate equation numbering
```

In [2]:
```
%%javascript
MathJax.Hub.Config({
    TeX: { equationNumbers: { autoNumber: "AMS" } }
});
```

In [5]:
```
%%javascript
MathJax.Hub.Queue(
  ["resetEquationNumbers", MathJax.InputJax.TeX],
  ["PreProcess", MathJax.Hub],
  ["Reprocess", MathJax.Hub]
);
```

Theoretical Model

Analogous to Contingent Valuation methods, Choice experiments also build on Random Utility Modeling (RUM), in that the starting point of the theoretical model is a specification of an indirect utility function (IUF) for a stipulated choice option. The main difference to CV is that the individual now faces **three or more simultaneous options** to choose from at each choice occasion. Furthermore, CE's typically offer more than one choice set to each survey respondent (at the risk, of course, of triggering ordering and sequencing effects).

Some upfront definitions:

- *Choice option or "alternative"*: A specific combiniation of attribute levels and price
- *Choice set*: A bundle of 3 or more options to choose from (typically 3, with one being the Status Quo (SQ))
- *Choice block*: A group of several choice sets to be considered sequentially by the respondent (typically 4-8 are given in a survey version)
- *Choice occasion*: A specific choice set within the choice block

Formally, let the Indirect Utility Function (IUF) for person i, choice occasion t, and alternative j be given as:

$$\begin{aligned} &U_{itj} = \mathbf{z}'_{itj}\boldsymbol{\theta} + \lambda\left(m_i - P_{itj}\right) + \epsilon_{itj}, \quad \text{with} \\ &\epsilon_{itj} \sim EV\left(0,1\right), \end{aligned} \tag{1}$$

where vector $\mathbf{z}_{itj}$ comprises the underlying attributes of a specific choice option, m_i denotes (typically annual) income, P_{itj} is the price or "bid" associated with the choice option, and ϵ_{itj} captures all other components that affect utility, but are not visible to the analyst. The error term in the basic RUM model is stipulated to follow a "Type-I Extreme Value (EV)" distribution with zero mean and unity scale.

As for the CV case, $P_{itj} = 0$ for the status quo option. Typically, the SQ option comprises the same attribute set as the policy options, but at different levels that remain unchanged over individuals. In some cases, as in the Red Tide example we'll study in some detail, there are no matching SQ attributes that make sense, in which case SQ utility is simply given as $U_{it0} = \lambda m_i + \epsilon_{it0}$.

We could now again divide by the marginal utility of income, λ to convert the model to "WTP-space." However, in this case there are no compelling reasons for this, such as econometric efficiency gains. We thus remain in "utility-space."

Respondent i will choose alternative j if it provides larger utility than all other options, that is:

$$prob\left(y_{itj} = j\right) = prob\left(U_{itj} - U_{ith} > 0\right), \ \forall h \neq j, \tag{2}$$

where y_{itj} is a binary indicator that takes the value of 1 if i chooses j on the t^{th} occasion, and a value of zero otherwise. Note that this does not simply break into independent products of binary decisions, since the error term of the winning utility becomes part of the differenced error term of *all utility differences*, and thus links the entire system of differenced utilities.

Instead, under the maintained distributional assumptions of the error term in (???) the probability of i selecting option j on occasion t can be conveniently expressed as:

$$\begin{aligned} \text{prob}\left(y_{itj} = 1\right) &= \frac{exp\left(\mathbf{x}_{itj}'\boldsymbol{\beta}\right)}{\sum_{j=1}^{J} exp\left(\mathbf{x}_{itj}'\boldsymbol{\beta}\right)}, \quad \text{where} \\ \mathbf{x}_{itj} &= \left[\begin{array}{cc} \mathbf{z}_{itj}' & P_{itj} \end{array}\right]', \quad \text{and} \quad \boldsymbol{\beta} = \left[\begin{array}{cc} \boldsymbol{\theta}' & -\lambda \end{array}\right]', \end{aligned} \tag{3}$$

This is the famous **"Conditional Logit" (CL)** form derived my nobel laureat Daniel McFadden in what must be one of the most highly cited papers in all of economics, despite its rather obscure appearance as a chapter in an edited volume (McFadden, 1974). A further distinction of CL specifications is made based on the nature of the given alternatives. Specifically, choice options in a typical CE do not represent existing real-world alternatives such as transportation ("bus," "car," "bike,") or food choices ("beef," "pork," "chicken"), but rather hypothetical mixes of attributes that vary in composition across alternatives and individuals. As such, the CE case constitutes what is generally referred to as an **"unlabeled" choice experiment** (e.g. Holmes et al., 2017}.

An important econometric implication of an unlabeled experiment is that all choice alternatives share the same set of coefficients, as is evident from (???).

As is also evident from (???), we collect attribute vector $\mathbf{z}_{itj}$ and bid P_{ijt} into a single data vector $\mathbf{x}_{itj}$, and corresponding coefficients into single coefficient vector $\boldsymbol{\beta}$, for ease of notation. As shown in (???) we enter price as a positive term and envision the marginal utility of income λ entering the model with a negative sign. This is purely based on the analyst's preferences and can be changed, as long as the interpretation remains clear.

As a final note, in the econometric model $\mathbf{x}_{itj}$ will usually include a constant term, but only for the SQ alternative. This coefficient thus captures unobservable elements that may influence respondents' decison to vote for or against the status quo.

Linear and nonlinear attribute space

Consider a single attribute, say x that, in reality, represents a continuous variable. As is typical in most CEs, this attribute will have no more than 3-5 pre-set levels in the different choice menus. For example, in our red tide forecasting application, the attribute "forecast accuracy" (theoretically continuous between 0 and 100) had set levels of 50%, 75% and 100%.

The analyst now has the choice of estimating a single coefficient β_x for this attribute, essentially forcing it to have a **linear effect** on WTP. Continuing with our example, this would imply that a step from 50-75% accuracy increases WTP by the same amount as a step from 75%-100%.

However, this constraint is not necessary from an econometric perspective, and likely violated in practice. Instead, I recommend to always break attributes like that into its individual levels and estimate a separate parameter for each level (minus the omitted baseline level). In our application, we included a separate binary indicator for "accuracy=75%," and a second one for "accuracy=100%," and the estimates emerged as anything but similar, as you will see in the application module.

This step-wise treatment of (essentially) continuous attributes is often referred to as the **non-linear model**.

WTP predictions

For the Conditional Logit model with linear IUF as given in (???), the **marginal WTP** for a 1-unit change in a given attribute can be obtained as the simple ratio of the attribute's coefficient over the MUI, i.e. as $\frac{\beta_x}{\lambda}$.

In contrast, WTP (Compensating surplus), labeled as w_i, for an entire bundle of attribute settings, say $\mathbf{z}_p$, is obtained as usual by equating indirect utility for the SQ settings $\mathbf{z}_0$ at full income m_i with indirect utility associated with the policy bundle and reduced income $m_i - w_i$. This produces the following expression:

$$w_i|\mathbf{z}_p, \boldsymbol{\theta}, \lambda = \frac{1}{\lambda}\left((\mathbf{z}_p - \mathbf{z}_0)'\boldsymbol{\theta}\right) \tag{4}$$

If there are no specific SQ attribute settings, but only a SQ constant, WTP can then be instead derived as:

$$w_i|\mathbf{z}_p, \boldsymbol{\theta}, \lambda = \frac{1}{\lambda}\left(\mathbf{z}_p'\boldsymbol{\theta}_z - \theta_{SQ}\right). \tag{5}$$

where we have separated the full coefficient vector $\boldsymbol{\theta}$ into a sub-vector $\boldsymbol{\theta}_z$ that corresponds to policy attributes, and the SQ coefficient θ_{SQ}.

Econometric model

Likelihood function

The sample likelihood for $i = 1 \dots N$ independent individuals, each facing T *independent choice occasions* involving J alternatives, is given by

$$p\left(\mathbf{y}|\boldsymbol{\beta}, \mathbf{X}\right) = \prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{j=1}^{J}\left(\frac{exp\left(\mathbf{x}_{itj}'\boldsymbol{\beta}\right)}{\sum_{j=1}^{J} exp\left(\mathbf{x}_{itj}'\boldsymbol{\beta}\right)}\right)^{y_{itj}}. \tag{6}$$

Taking the product over all individuals and choice occasions naturally implies that each occasion was truly treated as independent from all others, as instructed in the survey. We can test this assumption by comparing results from a "first-choice-only" model to the full-sample model, as you will do in PS 3.

Priors, posterior, and data augmentation

Adding the typical multivariate normal prior for $\boldsymbol{\beta}$, as we did for all previous models, we obtain the following joint posterior kernel:

$$p\left(\boldsymbol{\beta}|\mathbf{y},\mathbf{X}\right) \propto exp\left(-\frac{1}{2}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_0\right)'\mathbf{V}_0^{-1}\left(\boldsymbol{\beta}-\boldsymbol{\mu}_0\right)\right)\prod_{i=1}^{N}\prod_{t=1}^{T}\prod_{j=1}^{J}\left(\frac{exp\left(\mathbf{x}_{itj}'\boldsymbol{\beta}\right)}{\sum_{j=1}^{J}exp\left(\mathbf{x}_{itj}'\boldsymbol{\beta}\right)}\right)^{y_{itj}} \tag{7}$$

This expression does not have a well-understood statistical form, nor can it be broken into conditionals to set up a standard Gibbs Sampler. This is a typical situation where a Metropolis-Hastings (MH) approach can be helpful.

The Metropolis-Hastings algorithm

In essence, the MH algorithm is used to draw from an unknown distribution. This unknown "target distribution" can be the full posterior (as in our current case), or a conditional posterior. Either way, the general methodology is the same.

Let's assume the unknown posterior for some parameter θ (scalar or vector) is generically given by

$$p\left(\theta|\boldsymbol{\Gamma},\mathbf{y}\right) = \frac{p\left(\theta,\mathbf{y}|\boldsymbol{\Gamma}\right)}{p\left(\mathbf{y}|\boldsymbol{\Gamma}\right)} = \frac{p\left(\theta\right)p\left(\mathbf{y}|\theta,\boldsymbol{\Gamma}\right)}{p\left(\mathbf{y}|\boldsymbol{\Gamma}\right)}, \tag{8}$$

where $\boldsymbol{\Gamma}$ represents data $\mathbf{X}$ and potentially other model parameter. If $\boldsymbol{\Gamma}=\mathbf{X}$, then we are aiming to draw from the full posterior. If $\boldsymbol{\Gamma}$ contains other parameters, in addition to $\mathbf{X}$, we are drawing from a conditional. I will continue with the more general case where $\boldsymbol{\Gamma}$ can include both - it makes no difference for what follows.

We generally do know the mathematical form of the posterior kernel $\tilde{p}\left(\theta|\boldsymbol{\Gamma},\mathbf{y}\right) = p\left(\theta\right)p\left(\mathbf{y}|\theta,\boldsymbol{\Gamma}\right)$ (even though we don't kow its statistical "family"). The unknown element is the *normalizing constant* $p\left(\mathbf{y}|\boldsymbol{\Gamma}\right)$. Equation (7) is an example of $\tilde{p}\left(\theta|\boldsymbol{\Gamma},\mathbf{y}\right)$.

The rationale of the MH algorithm is to use $\tilde{p}\left(\theta|\boldsymbol{\Gamma},\mathbf{y}\right)$, plus a **candidate-generating density (CGD)**, often also referred to as **proposal density** $q\left(\theta\right)$ to obtain draws from the unknown $p\left(\theta|\boldsymbol{\Gamma},\mathbf{y}\right)$. Note that $q\left(\theta\right)$ can also be a function of the data and / or other prameters in addition to θ.

Suppose that the most recent draw of θ in the GS (which will initially be the starting draw) is θ^a. Now obtain a new "candidate draw" of θ, call it θ^b, from $q\left(\theta\right)$. The new draw of θ is then accepted with probability

$$\alpha\left(\theta^a,\theta^b\right) = min\left(\frac{p\left(\theta^b|\boldsymbol{\Gamma},\mathbf{y}\right)q\left(\theta^a\right)}{p\left(\theta^a|\boldsymbol{\Gamma},\mathbf{y}\right)q\left(\theta^b\right)},1\right) = min\left(\frac{\tilde{p}\left(\theta^b|\boldsymbol{\Gamma},\mathbf{y}\right)q\left(\theta^a\right)}{\tilde{p}\left(\theta^a|\boldsymbol{\Gamma},\mathbf{y}\right)q\left(\theta^b\right)},1\right) \tag{9}$$

since the normalizing constant $p\left(\mathbf{y}|\boldsymbol{\Gamma}\right)$ cancels out in the ratio. We usually work with logs:

$$\begin{aligned} log\left(\alpha\left(\theta^a,\theta^b\right)\right) =& min\left(log\,\tilde{p}\left(\theta^b|\boldsymbol{\Gamma},\mathbf{y}\right) + log\,q\left(\theta^a\right) - log\,\tilde{p}\left(\theta^a|\boldsymbol{\Gamma},\mathbf{y}\right) - log\,q\left(\theta^b\right),0\right) = \\ & min\left(\left(log\,\tilde{p}\left(\theta^b|\boldsymbol{\Gamma},\mathbf{y}\right) - log\,q\left(\theta^b\right)\right) - \left(log\,\tilde{p}\left(\theta^a|\boldsymbol{\Gamma},\mathbf{y}\right) - log\,q\left(\theta^a\right)\right),0\right) \end{aligned} \tag{10}$$

In practice this is implemented by comparing the α value to a random uniform [0,1] draw (or, equivalently, $log\,\alpha$ to the log of a uniform draw). If α exceeds the random value, the new draw θ^b is accepted, otherwise θ^a remains the most current draw. As discussed in Gelman et al (Ch. 12) , Koop (Ch. 5) and KPT, Ch. 11, after a sufficient number of "burn-ins" the sequence of draws of θ will converge to the desired underlying posterior density.

Also note that the **generic GS is a special case of the MH** with $q\left(\theta^j\right) = \tilde{p}\left(\theta^j|\boldsymbol{\Gamma},\mathbf{y}\right),\ j=a,b$ such that the acceptance probability is always one, i.e. every new draw of θ is accepted by default.

MH for the Conditional Logit model

There are many possible proposal densities for the MH approach. Here we will use one that produces draws that are highly efficient, i.e. not highly correlated, as is always desirable in Bayesian estimation. It is a type of "*Independence Chain*" (IC) MH, where the current draw of θ is *not* a direct moment (e.g. mean) of the proposal function. This lessens autocorrelation in the chain of draws.

The specific IC-MH version we will use in this application was proposed by Rossi et al. (2005) for the Conditional Logit model, though it is applicable to a wide range of other specifications. It works as follows:

1) At each iteration of the posterior sampler find the mode of the posterior kernel $\tilde{p}\left(\boldsymbol{\beta}|\mathbf{y},\mathbf{X}\right)$, i.e. use a short Maximum Likelihood (MLE) routine to find the $\boldsymbol{\beta}$ that maximizes this function. Let's call it $\tilde{\boldsymbol{\beta}}$. Furthermore, let the Hessian (matrix of second derivatives) coming out of this optimization be labeled $\tilde{H}$.The MLE step can be implemented with analytical gradient and Hessian, and is thus both fast and precise. Details on these MLE components are given in Moeltner et al. (2021).

2) Draw a candidate $\boldsymbol{\beta}^c$ from $q\left(\boldsymbol{\beta}^c|\tilde{\boldsymbol{\beta}},\tau * \left(-\tilde{H}\right)^{-1},\nu\right)$, where $q\left(.\right)$ is the multivariate t-distribution with mean $\tilde{\boldsymbol{\beta}}$, scale matrix $\tau * \left(-\tilde{H}\right)^{-1}$, and degrees of freedom ν. The tuning scalar τ and degress of freedom parameter ν are chosen at the onset. We will come back to these shortly.

3) The new draw is accepted over the old draw $\boldsymbol{\beta}^0$ with probability:

$$\alpha\left(\boldsymbol{\beta}^0,\boldsymbol{\beta}^c\right) = \frac{\tilde{p}\left(\boldsymbol{\beta}^c|\mathbf{y},\mathbf{X}\right) * q\left(\boldsymbol{\beta}^0|\tilde{\boldsymbol{\beta}},\tau * \left(-H\right)^{-1},\nu\right)}{\tilde{p}\left(\boldsymbol{\beta}^0|\mathbf{y},\mathbf{X}\right) * q\left(\boldsymbol{\beta}^c|\tilde{\boldsymbol{\beta}},\tau * \left(-H\right)^{-1},\nu\right)} \tag{11}$$

or, in log form:

$$log\left(\alpha\left(\boldsymbol{\beta}^0,\boldsymbol{\beta}^c\right)\right) = min\left(\left(log\,\tilde{p}\left(\boldsymbol{\beta}^c|\mathbf{y},\mathbf{X}\right) - log\,q\left(\boldsymbol{\beta}^c|.\right)\right) - \left(log\,\tilde{p}\left(\boldsymbol{\beta}^0|\mathbf{y},\mathbf{X}\right) - log\,q\left(\boldsymbol{\beta}^0|.\right)\right),0\right), \tag{12}$$

where $\tilde{p}\left(.\right)$ is given in (7).

What is a desirable acceptance rate? For the GS, every draw is accepted by default, as noted above. For a MH with a (typically) high degree of autocorrelation (e.g. when the current draw is used as the mean of the CGD), accepting too many draws would imply being "stuck" in one corner of the posterior, without enough "jumping around." Accepting too little will equally imply being stuck at a single point for long stretches of draws. In those cases, one typicaly aims for acceptance rates of 0.25 -0.45.

However, for the IC-MH, where draws are relatively more independent, we would like to accept a high proportion, say in the 70-80% range. This is accomplished by setting tuners τ and ν accordingly in a series of trial runs with, say, 500-1000 iterations.

Typically, the acceptance rate AR (= proportion of accepted draws out of the set of keeper draws) is reported in the final output along with the posterior results for $\boldsymbol{\beta}$.

After obtaining these draws from the posterior sampler, inference can proceed as usual by inspecting the posterior distributions of individual coefficients, computing HPDI bounds, and generating posterior predictive distibutions for WTP scenarios.

References:

McFadden, D. 1974. "Conditional logit analysis of discrete choice behavior," in P. Zarembka, ed. *Frontiers of Econometrics*. New York: Academic Press.

Holmes, T., Adamowicz, W., Carlsson, F., 2017. "Choice experiments," in: Champ, P., Boyle, K., Brown, T. (Eds.), *A primer on nonmarket valuation*. Springer, pp. 133-186.

Moeltner, K., T. Fanara, H. Foroutan, R. Hanlon, V. Lovko, S. Ross, and D. Schmale III, "Harmful algal blooms and toxic air: The economic value of improved forecasts," paper presented at the annual meetings of the European Association of Environmental and Resource Economists (EAERE), virtual, Jun. 25, 2021.

Rossi, P., Allenby, G., McCulloch, R., 2005. *Bayesian Statistics and Marketing*. John Wiley & Sons.