Mixlelast: Stata Module for Mixed Logit Elasticities and Marginal Effects

Lars Zeigermann
lars.zeigermann@posteo.de

May 27, 2024

Abstract

Thanks to their greater flexibility and more realistic substitution patterns compared to simpler discrete choice models, mixed logit models are very popular in discrete choice analysis. While fitting a mixed logit model in Stata using `mixlogit` (Hole 2007) is straight-forward, calculating elasticities and marginal effects is not. This article describes `mixlelast`, a post-estimation command for `mixlogit`. It allows the researcher to compute various forms of mixed logit sample elasticities and marginal effects and to obtain bootstrapped standard errors and confidence intervals.

1 Introduction

Mixed logit models have become very popular in discrete choice analysis. This is due to their greater flexibility and the more realistic substitution patterns compared to simpler discrete choice models. Thanks to the excellent user-written `mixlogit` by Hole (2007), it is fairly simple for the researcher to fit mixed logit models in Stata. As there exists no straight-forward interpretation for the estimated parameters beyond theirs signs, researchers often compute elasticities or marginal effects. In contrast to simpler logit models, computing elasticities and marginal effects for mixed logit models is not trivial and requires simulation. This article describes `mixlelast`, a post-estimation command after `mixlogit`, which allows the user to obtain mixed logit elasticities and marginal effects of various flavours and to compute bootstrapped standard errors and confidence intervals.

The remainder of this paper is structured as follows: Section 2 briefly recapitulates the mixed logit model. The various sorts of mixed logit elasticities and marginal effects are discussed in sections 3 and 4. Section 5 deals with sample elasticities and marginal effects in the case of heterogeneous choice sets. A bootstrap procedure for standard errors and confidence intervals is

I am grateful to Anna Lu and Arne Risa Hole for their very helpful comments and suggestions. I would also like to thank Tomaso Duso, Florian Heiss, André Romahn, Hannes Ullrich and the participants of the DICE research workshop at HHU Düsseldorf for their feedback. The code of `mixlelast` builds upon `mixlpred` written by Arne Risa Hole.

presented in section 6. Finally, the syntax of `mixlelast` is described in section 7 and section 8 provides some examples.

2 Mixed logit model[1]

As in Revelt and Train (1998) I assume N decision makers to choose among J alternatives on T choice occasions. The utility for individual n from choosing alternative j at period t is given by

$$U_{njt} = \beta_n' x_{njt} + \varepsilon_{njt}$$

where β_n is a vector of individual-specific coefficients for decision maker n and x_{njt} is a vector of attributes. The error term, ε_{njt}, is assumed to be independently and identically distributed extreme value I. The density of β is denoted $f(\beta|\theta)$, where θ are the parameters of the distribution, i.e. the mean and the (co)variance.

The conditional probability, i.e. conditional on β_n, for individual n of choosing alternative i on choice occasion t is

$$L_{nit}(\beta_n) = \frac{exp(\beta_n' x_{nit})}{\sum_{j=1}^{J} exp(\beta_n' x_{njt})}. \quad (1)$$

Then the conditional probability of the observed sequences of choices made by decision maker n is given by

$$S_n(\beta_n) = \prod_{t=1}^{T} L_{ni(n,t)t}(\beta_n)$$

where $i(n,t)$ denotes the chosen alternative. The *unconditional* probability of the observed sequence of choices is the conditional probability integrated over the distribution of β:

$$P_n(\theta) = \int S_n(\beta|\theta) f(\beta) d\beta$$

The researcher specifies the form of the density function f and estimates its parameters θ by maximum likelihood. The log likelihood function is given by

$$LL(\theta) = \sum_{n=1}^{N} \ln P_n(\theta). \quad (2)$$

There exists no analytical solution for this expression and it needs to be approximated using simulation techniques, see Train (2009). The simulated equivalent of (2) is given by

$$SLL(\theta) = \sum_{n=1}^{N} \ln \left[\frac{1}{R} \sum_{r=1}^{R} S_n(\beta^r) \right]$$

where R is the number of draws and β^r is the rth draw from $f(\beta|\theta)$.

[1] This section draws heavily from Hole (2007). See Revelt and Train (1998) and Train (2009) for an in-depth discussion of the model.

The estimated parameters $\hat{\theta}$ have, beyond their sign, no straight-forward interpretation. To be able to provide meaningful and interpretable results for discrete choice models, researcher often compute elasticities or marginal effects of various kinds.

3 Elasticities

Elasticities are a unitless measure which describes the relationship between a percentage change in an attribute of an alternative and the percentage change in the choice probability, ceteris paribus, i.e. everything else remaining unchanged. They can take various forms, which can be distinguished along three categories, see Hensher et al. (2015).

First, there exist a distinction between *direct* and *cross* elasticities. The former measure the percentage change in the choice probability of an alternative with respect to a change in an attribute of the *same* alternative. A prominent example for direct elasticities are own-price elasticities, i.e. the change in the choice probability of a product when its own prices changes. The latter, in contrast, describes the change in the choice probability of one alternative with respect to a percentage change in an attribute of a *competing* alternative. Here cross-price elasticities are a common example describing the change in the choice probability resulting from a price change of a rival product.

Second, we differentiate between two methods of calculation, namely the *point* and the *arc* method. Point elasticities use differentiation and evaluate a marginal change in an attribute of one alternative. The arc method calculates the choice probabilities before and after a (non-marginal) attribute change and evaluates the difference. This allows the researcher to compute the effects of a discrete change in an attribute, e.g. a price increase of 20 per cent.

Third, there is a distinction between *individual* and *sample* elasticities. Data on individual choices of decision makers allows to compute individual elasticities for every alternative on every choice occasion. However, one usually ends up with a very large number of elasticities and researchers are often more interested in average elasticities over the entire sample. As sample elasticities are unweighted or weighted averages of individual elasticities, individual elasticities will be discussed first before turning to different ways of aggregation.

3.1 Individual elasticities

3.1.1 Point elasticities

Point elasticities describe the percentage change in the choice probability given a small change in one attribute. For calculation we use differentiation. Recalling the s-shape of the logit curve, it becomes obvious why point elasticities are only valid for marginal changes at a particular point on the logit curve. With a non-marginal change in an attribute, we move along the logit curve and might find a considerably different slope compared to the original attribute value. Ignoring the impact of the non-linearity of the logit curve can give rise to substantially biased results.

The researcher should note that calculating point elasticities does only make sense for continuous variables. Although it is mathematically possible to compute point elasticities for integer or dummy variables, it is meaningless. The appropriate way to evaluate the impact of a discrete change in an attribute is to use the arc method as described later.

Direct elasticities Stated formally, the direct point elasticity of alternative i of decision maker n for a marginal change in the mth attribute is given by[2]

$$E_{nix_{ni}^m} = -\frac{x_{ni}^m}{P_{ni}} \frac{\partial P_{ni}}{\partial x_{ni}^m} \tag{3}$$

where

$$P_{ni} = \int L_{ni}(\beta) f(\beta) d\beta. \tag{4}$$

Recall that L_{ni} is the conditional probability of decision maker n of choosing alternative i as given in (1). Taking the derivative of (4) with respect to x_{ni}^m and plugging into (3) yields

$$E_{nix_{ni}^m} = -\frac{x_{ni}^m}{P_{ni}} \int \beta^m L_{ni}(\beta)(1 - L_{ni}(\beta)) f(\beta) d\beta \tag{5}$$

where β^m is the mth element of vector β.

The integral in (5) cannot be solved analytically and is therefore approximated using simulation techniques. The simulated version of (5) is given by

$$E_{nix_{nj}^m} \approx -\frac{x_{nj}^m}{P_{ni}} \left(\frac{1}{R} \sum_{r=1}^{R} \beta^{mr} L_{ni}(\beta^r)(1 - L_{ni}(\beta^r)) \right)$$

where β^r denotes the rth draw from the distribution of β, see Train (2009).

Cross elasticities The cross point elasticity of alternative i for individual n given a marginal change in the mth attribute of the rival alternative j is then

$$E_{nix_{nj}^m} = \frac{x_{nj}^m}{P_{ni}} \frac{\partial P_{ni}}{\partial x_{nj}^m}.$$

Plugging in the first derivative of (4) with respect to x_{nj}^m yields

$$E_{nix_{nj}^m} = \frac{x_{nj}^m}{P_{ni}} \int \beta^m L_{ni}(\beta) L_{nj}(\beta) f(\beta) d\beta. \tag{6}$$

The simulated equivalent of (6) is then given by

$$E_{nix_{nj}^m} \approx \frac{x_{nj}^m}{P_{ni}} \left(\frac{1}{R} \sum_{r=1}^{R} \beta^{mr} L_{ni}(\beta^r) L_{nj}(\beta^r) \right).$$

3.1.2 Arc elasticities

In contrast to point elasticities, arc elasticities consider a discrete change in one attribute rather than a marginal change.[3] Instead of differentiation, we calculate expected probabilities for the

[2] For ease of notation, I drop T in what follows. I.e. I assume each of the N decision makers to choose only once.
[3] The term arc elasticity is not used consistently in the literature. Sometimes researchers refer to arc elasticities when the first derivative is not multiplied by x/P but by the average over the before and after values, i.e. $\Delta x/\Delta P$. This approach can be seen as an linear approximation which is ignoring the non-linearity of the logit curve.

original attribute value and with a changed attribute value and evaluate the difference. This approach handles the non-linearity of the logit curve which affects the results when non-marginal attribute changes are considered.

Direct elasticities Formally, the direct arc elasticity of alternative i for individual n given a discrete change in the mth attribute is given by

$$E_{nix_{ni}^m} = \frac{\bar{x}_{ni}^m}{\bar{P}_{ni}} \frac{\Delta P_{ni}}{\Delta x_{ni}^m}. \tag{7}$$

First consider the second right-hand side term of (7). The numerator, ΔP_{ni}, is the difference in the choice probability for the attribute value before the change, x_{ni}, and after the change $\hat{x}_{ni}$, i.e.:

$$\Delta P = P_{ni}(x_{nj}^m) - P_{ni}(\hat{x}_{nj}^m)$$

The denominator Δx_{ni}^m is simply the difference between the original and changed attribute value:

$$\Delta x_{nj} = x_{nj}^m - \hat{x}_{nj}^m$$

The first right-hand side term comprises the average choice probability before and after the attribute change[4]

$$\bar{P} = (P_{ni}(x_{nj}^m) + P_{ni}(\hat{x}_{nj}^m))/2$$

and the attribute mean

$$\bar{x} = (x_{nj}^m + \hat{x}_{nj}^m)/2.$$

For integer variables, we can now specify $\hat{x} = x + a$, where a can take any integer value. By the same token, elasticities for dummy variables are calculated by setting $\hat{x} = 0$ if $x = 1$ and vice versa.

Cross elasticities Equivalently, cross arc elasticities are given by

$$E_{nix_{nj}^m} = \frac{\bar{x}_{nj}^m}{\bar{P}_{ni}} \frac{\Delta P_{ni}}{\Delta x_{nj}^m}$$

where ΔP_{ni} now denotes the difference in the choice probability for alternative i with the original attribute of j, x_{nj}^m and the changed attribute, $\hat{x}_{nj}^m$.

3.1.3 Arc vs. point elasticities

Having established the two different methods, the question arises, which is the correct one to use. For dummy and integer variables, the answer is obvious as marginal changes are meaningless. Here the researcher should always use arc elasticities. For continuous variables, the answer depends strongly on the question the researcher seeks to answer. If the researcher wants to evaluate the decision makers sensitivity to attribute changes and establish general substitution patterns across

[4] The approach followed here, i.e. taking the mean probability and mean attribute value, is sometimes referred to as the midpoint method. Two alternatives not considered multiply by either the original or the changed values and the respective predicted probabilities.

alternatives, the point method is appropriate. If, in contrast, we want to evaluate how a substantial change in an attribute, e.g. a price increase by 20 percent, effects choice behaviour, we need to obtain arc elasticities.

Calculating arc elasticities for very small discrete attribute changes should generate results very close to point elasticities. Without any post-estimation command available, this was a practicable solution to obtain approximated point elasticities.[5] With `mixlelast`, the researcher is strongly advised to use the point method rather than calculating arc elasticities for very small changes. First, arc elasticities for small changes are only an approximation of point elasticities. Second, calculating arc elasticities is more computationally intensive which may lead to a significantly longer computation time for large data sets.

3.1.4 Interpretation

Elasticities can take any value between infinity and minus infinity. Positive direct elasticities indicate that an increase in an attribute of an alternative leads to an increase in the choice probability of that alternative. For cross elasticities, positive values indicate that an increase in an attribute leads to a higher choice probability in the rival alternative. For negative elasticities, the effects are reversed.

If a one percent change in an attribute results in a one percent change in the choice probability, we speak about unit elasticity. If the percentage in the choice probability is smaller or larger than one, we speak about relatively inelastic and relatively elastic elasticities respectively. Two extreme cases are perfectly inelastic and perfectly elastic elasticities. In the former case, an attribute change leaves the choice probability unchanged. In the latter case, the choice probability will drop to zero following an increase of one percent in the attribute. The various types of elasticities are summarized in table 1 below.

	Absolute value of elasticity	**Direct elasticity**	**Cross elasticity**
Perfectly inelastic	$E = 0$	no change in P_i	no change in P_j
Relatively inelastic	$0 < E < 1$	change in P_i less than one percent	change in P_j less than one percent
Unit elastic	$E = 1$	no change in P_i	no change in P_j
Relatively elastic	$1 < E < \infty$	change in P_i larger than one percent	change in P_j larger than one percent
Perfectly elastic	$E = \infty$	change in P_i is ∞	change in P_j is ∞

Table 1: Changes in the choice probability given a one percent change in an attribute of alternative i, adapted from Hensher et al. (2015)

[5] Cameron and Trivedi (2010) advocate calculating arc elasticities using a discrete change of a one thousandth of the standard deviation of the attribute to approximate point elasticities.

3.1.5 Mixed vs. conditional logit elasticities

Having established expressions for individual direct and cross point elasticities, we can now compare mixed logit elasticities with those from standard conditional logit models.

The conditional logit formula (McFadden 1973) is

$$P_{nit} = \frac{exp(\beta' x_{nit})}{\sum_{j=1}^{J} exp(\beta' x_{njt})}.$$

In contrast to the mixed logit model, the β coefficients are assumed to be identical for all decision makers. Now, direct elasticities are given by

$$E_{nix_{ni}^m} = -\frac{x_{ni}^m}{P_{ni}} \frac{\partial P_{ni}}{\partial x_{ni}^m}. \tag{8}$$

Taking the derivative in (8) yields

$$E_{nix_{ni}^m} = -\frac{x_{ni}}{P_{ni}} \beta P_{ni}(1 - P_{ni})$$

which simplifies to

$$E_{nix_{ni}^m} = -x_{ni}\beta(1 - P_{ni}). \tag{9}$$

Contrary to the complex mixed logit elasticity in (5) which contains integrals which have to be simulated, direct logit elasticities given in (9) are a simple function of the attribute value, the estimated coefficient and the predicted choice probability.

More insightful, however, is the comparison of cross elasticities. Conditional logit cross elasticities are given by

$$\begin{aligned} E_{nix_{nj}^m} &= \frac{x_{nj}^m}{P_{ni}} \frac{\partial P_{ni}}{\partial x_{nj}^m} \\ &= \frac{x_{nj}^m}{P_{ni}} \beta P_{ni} P_{nj}. \end{aligned}$$

As before, P_{ni} cancels out:

$$E_{nix_{nj}^m} = x_{nj}^m \beta P_{nj} \tag{10}$$

As we can see in (10), cross elasticities are identical for all i. Hence, a change in an attribute of j gives rise to the exact same change in the choice probability of all rival alternatives. This property of logit elasticities is sometimes referred to as "proportionate shifting".[6] While this might be realistic for some choice situations, for many it is not. If some alternatives resemble each other closely, while others are very different, we would expect substitution between the former to be stronger than between the latter.

Here, the strength of mixed logit models comes into play. Recall the mixed logit cross elasticities

[6]See Train (2009) for an in-depth discussion of the properties of logit elasticities.

given in (6):

$$E_{nix^m_{nj}} = -\frac{x^m_{nj}}{P_{ni}} \int \beta^m L_{ni}(\beta) L_{nj}(\beta) f(\beta) d\beta$$

This expression is different for each alternative i. The percentage change in the probability P_{ni} depends on the correlation of L_{ni} and L_{nj} over different realisations of β. For alternatives between which x^m is positively correlated, substitution will be stronger than for alternatives with no or negative correlation, see Train (2009). Therefore, similar alternatives will exhibit much stronger substitution than those which are very different.

The comparison also illustrates that using the more flexible mixed logit framework does not come without costs. While there are simple closed form solutions for the logit model and the resulting elasticities, we need to employ simulation techniques both for estimation and calculation of elasticities for the mixed logit model.

3.2 Sample elasticities

The elasticities above are calculated for one specific decision maker in one particular choice situation. With J alternatives this gives J^2 values on a single choice occasion; J direct elasticities and $(J-1) \cdot J$ cross elasticities. For N choice occasion, this amounts to $N \cdot J^2$ values.

Thus, the first obvious reason why researchers usually want to compute sample elasticities, is to reduce the amount of output produced and to make it readable. Taking the sample averages reduces the number of values obtained by a factor of N to J^2. These can be practically displayed in a JxJ matrix with direct elasticities on the diagonal. A second and equally important argument in favour of sample elasticities is that the underlying model is estimated on a sample of choice data. Meaningful elasticities can therefore only be calculated for the sample, but not separately for each decision maker, see Hensher et al. (2015).

3.2.1 Unweighted sample average

The most straight-forward way of obtaining samples averages is to simply calculate the mean over all decision makers in the sample. Sample direct and cross elasticities are then given by

$$SE_{ix^m_i} = \frac{1}{N} \sum_{n=1}^{N} E_{nix^m_i}$$

and

$$SE_{ix^m_j} = \frac{1}{N} \sum_{n=1}^{N} E_{nix^m_j}$$

respectively, where $E_{ix^m_i}$ and $E_{ix^m_j}$ can either be individual point or arc elasticities.

3.2.2 Probability weighted sample average

In the simple approach above, all elasticities enter with the same weight, regardless of the contribution to the choice outcome of each alternative. Suppose an alternative has a very high choice probability in one choice occasion but a low one in another. Simply averaging over the elasticities from those two choice occasions ignores this fact and is sometimes referred to as "naive pooling", see Louviere et al. (2000). To account for this, we can compute probability weighted sample averages using the decision maker's choice probabilities as weights[7]

Formally, probability weighted sample elasticities are then

$$SE_{ix_i^m} = \left(\sum_{n=1}^{N} P_{ni} E_{nix_i^m} \right) \Big/ \sum_{n=1}^{N} P_{ni}$$

and

$$SE_{ix_j^m} = \left(\sum_{n=1}^{N} P_{ni} E_{nix_j^m} \right) \Big/ \sum_{n=1}^{N} P_{ni}$$

where, once again, $E_{ix_i^m}$ and $E_{ix_j^m}$ can be calculated using the point or the arc method.

4 Marginal effects

In contrast to elasticities, which are we have discussed in section 3, marginal effects describe a change in the choice probability given a *unit* change in an attribute.

The direct marginal effect for decision maker n and alternative i with respect to a unit change in x_{ni}^m is given by

$$ME_{nix_{ni}^m} = \frac{\partial P_{ni}}{\partial x_{ni}^m}. \tag{11}$$

Comparing (11) and the expression for direct direct point elasticities given in (3), the difference between the two becomes obvious. It is the second term in (7), $\frac{x_{ni}^m}{P_{ni}}$, which transforms a marginal effect into an elasticity.

Equivalently, the cross marginal effect is given by

$$ME_{nix_{nj}^m} = \frac{\partial P_{ni}}{\partial x_{nj}^m}.$$

For the same reason as for elasticities we need to consider discrete changes for discrete and dummy variables.[8] Direct and cross marginal effects for discrete and dummy variables using the arc method are given by

$$ME_{nix_{ni}^m} = \frac{\Delta P_{ni}}{\Delta x_{ni}}$$

[7] For a discussion of these issues (in the context of the conditional logit model) see Louviere et al. (2000).

[8] Speaking about marginal effects in this context falsely suggests marginal changes which is why researchers sometimes refer to incremental effects.

and

$$ME_{nix^m_{nj}} = \frac{\Delta P_{ni}}{\Delta x^m_{nj}}.$$

The same methods of aggregation discussed before can be applied.

5 Heterogeneous choice sets

So far I have implicitly assumed choice sets to be homogeneous across individuals, i.e. $J^n = J$. With homogeneous choice sets we get a full JxJ matrix of direct and cross effects for all N choice occasions. However, it is possible that the availability of alternatives differs across individuals.[9] That is only a subset J^n of the J alternatives is available to decision maker n. Then cross and directs effects are only defined for the alternatives in J^n. That is, we get J^n direct effects and $(J^n - 1) * J^n$ cross effects between all available alternatives. Considering the entire sample of all N decision makers, we obtain between one and N individual direct effects for an alternative: N if the respective alternative is contained in all choice sets and one if it is available on only a single choice occasion. For cross effects between two alternatives to be defined, both need to be available at the same time. This can be the case at all N choice occasions, only a subset of N or even at zero choice occasions.

Now, the question arises how should we calculate averages over the sample of decision makers which are heterogeneous regarding the alternatives to choose from. In the following, I will first discussion aggregation of direct effects before turning to cross effects.

5.1 Direct Elasticities

Calculating sample averages for direct effects, we generally have two possibilities. First, we consider all N decision makers in the sample no matter if the respective alternative is available or not. That is, we also include all those who do not at all react to any attribute change and set their individual elasticities to zero. Let's denote this **type I** aggregation. Secondly, we restrict our attention to those who can actually choose the alternative and will hence be affected by a change. Let's call this aggregation of **type II**.

Unweighted sample averages Formally, unweighted sample averages of type I are given by

$$SE^I_{ii} = \frac{1}{N} \sum_{n=1}^{N} E_{nii}.$$

Type II unweighted sample elasticities are then

$$SE^{II}_{ii} = \frac{1}{N^i} \sum_{n \in N^i} E_{nii}$$

where N^i denotes the subset of decision makers which hold alternative i in their choice set.

[9] The availability of alternatives might even differ across choice occasions faced by the same decision maker, if we deal with panel data where individuals choose repeatedly.

Probability weighted sample averages Applying the same logic of aggregation to probability weighted sample elasticities, we get the the following expression for type I:

$$SE_{ii}^{I} = \sum_{n=1}^{N} P_{ni} E_{nii} \bigg/ \sum_{n=1}^{N} P_{ni}$$

The observant reader will have noticed that the choice probability P_{ni} is not defined if alternative i is not in n's choice set. For convenience, we set it to zero indicating that alternative i is chosen with zero probability. For this reason, type I weighted sample averages are equivalent to their type II counterparts given by

$$SE_{ii}^{II} = \sum_{n \in N^i} P_{ni} E_{nii} \bigg/ \sum_{n \in N^i} P_{ni}.$$

5.2 Cross elasticities

There exist three types of aggregation for cross elasticities which I denote **type I**, **IIa** and **IIb**. First, **type I** aggregation considers all decision makers no matter if either i or j is present in the choice set. Second, we can aggregate over those decision makers which hold i in their choice sets no matter if alternative j is also present or not. Let's denote this **type IIa**. Third, we can restrict our attention to those decision makers which have both i and j in their choice sets, which we denote **type IIb** aggregation.

Unweighted sample averages Formally, unweighted average sample elasticities of type I are given by

$$SE_{ij}^{I} = \frac{1}{N} \sum_{n=1}^{N} E_{nij}.$$

Aggregation by type IIa and IIb yields

$$SE_{ij}^{IIa} = \frac{1}{N^i} \sum_{n \in N^i} E_{nij}$$

and

$$SE_{ij}^{IIb} = \frac{1}{N^{ij}} \sum_{n \in N^{ij}} E_{nij}$$

respectively, where N^{ij} denotes the subset of N where both alternatives are available.

Probability weighted sample averages Probability weighted sample elasticities constructed according to type I are then

$$SE_{ij}^{I} = \sum_{n=1}^{N} P_{ni} E_{nij} \bigg/ \sum_{n=1}^{N} P_{ni}.$$

Again, for all decision makers which did not have alternative i in their choice set, we set the probability and hence the weight to zero. For this reason, type I and type IIa aggregation given below are equivalent:

$$SE_{ij}^{IIa} = \sum_{n \in N^i} P_{ni} E_{nij} \bigg/ \sum_{n \in N^i} P_{ni}$$

Sample cross elasticities computed of type IIb are then given by

$$SE_{ij}^{IIb} = \sum_{n \in N^{ij}} P_{ni} E_{nij} \Big/ \sum_{n \in N^{ij}} P_{ni}.$$

5.3 Choice of aggregation type

Having established different ways of aggregation, we need to determine which is the correct one to use. Once again, the answer very much depends on the question we are asking. If we are interested in a measure of how the entire sample, being affected or not, reacts to changes, we want to employ type I. If, however, we care about those who are effected only, type II is what we need. For cross-effects we further need to decide if all individuals who hold i in their choice set are the sub-sample of interest (type IIa) or if we want to restrict our attention to those who can actually also choose j (type IIb).

6 Standard errors and confidence intervals

Recall that when estimating a mixed logit model, we in fact estimate the parameters of the distribution denoted θ, i.e. the mean and the (co)variance. In what we have discussed so far, we simply take the point estimates $\hat{\theta}$ of θ to simulate elasticities and marginal effects as described above. This gives us sample effects and standard deviations describing how much individual effects vary across the sample. However, we have ignored the sampling variance $\widehat{W}$ of the estimated $\hat{\theta}$. To obtain a measure of precision of our average sample effects, i.e. standard errors or confidence intervals, we need to incorporate the sampling variance of the estimated coefficients, $\widehat{W}$, into our calculation of elasticities and marginal effects.

We can generally think of three different methods for generating standard errors and confidence intervals, namely the non-parametric bootstrap, the parametric bootstrap (Krinsky-Robb method) and the delta-method. In the following we will consider the method proposed by Krinsky and Robb (1986, 1990) only as it does not require re-estimating the model and is fairly easy to implement.[10]

The Krinsky-Robb method involves the following four steps:

1. Compute the Cholesky-decomposition of the covariance matrix $\widehat{W}$, yielding a lower triangular matrix L such that $LL' = \widehat{W}$, see e.g. Greene (2008).

2. For $t = 1, \ldots, T$ create a $Kx1$ vector of random draws from a standard normal distribution, where K is the length of vector θ and label it η^t. Then create $\theta^t = \hat{\theta} + L\eta^t$.

3. For $t = 1, \ldots, T$ use θ^t to simulate elasticities or marginal effects as described above.

[10] For the non-parametric bootstrap we would have to re-estimate the model multiple times, which is often far too time-consuming and bears the risk that the model does not converge for some sub-samples. The delta method, in contrast, is relatively complex for mixed logit models and difficult to implement. A second downside of the delta-method is that confidence intervals are symmetric about the mean by construction.

4. Calculate the mean sample effect by averaging over the T draws. To obtain standard errors, compute the variance and take the square root. For the 95% confidence interval, take the 0.025 and 0.975 percentile.

7 mixlelast

Syntax

`mixlelast` [*if*] [*in*] , `alternatives(`*varname*`)` [`for(`*varname*`)` `marginaleffect` `percentchange(#)` `absolutechange(#)` `dummy` `weighted` `nosd` `nrep(#)` `burn(#)` `hettype()` `krobb(#)` `krse` `krlevel(#)` `kruserdraws` `krburn(#)` `quietly`]

Options

`alternatives(`*varname*`)` is required and identifies the alternatives. *varname* must be numeric.

`for(`*varname*`)` specifies the variable for which the elasticities/marginal effects are calculated. The default is the first variable specified in `mixlogit`'s `rand` option.

`percentchange(#)` allows the user to compute arc elasticities/marginal effects for an attribute change of # percent. This option cannot be combined with `absolutechange` and `dummy`.

`absolutechange(#)` allows the user to compute arc elasticities/marginal effects for an absolute attribute change of #. This option cannot be combined with `percentchange` and `dummy`.

`dummy` allows the user to compute arc elasticities/marginal effects for dummy variables. This option cannot be combined with `percentchange` and `absolutechange`.

`marginaleffect` allows the user to obtain marginal effects. The default is elasticities.

`weighted` specifies that probability weighted sample averages are computed. The default is unweighted sample averages.

`nosd` prevents the calculation and display of standard deviations.

`nrep(#)` specifies the number of Halton draws used for simulation. The default is the number specified for `mixlogit`, which in turn has a default of 50.

`burn(#)` specifies the number of initial sequence elements to drop when creating the Halton sequences. The default is the number specified for `mixlogit`, which in turn has a default of 15.

`hettype()` allows the user to specify the type of aggregation when choice sets are heterogenous. The default is I for unweighted and IIa for probability weighted sample averages. `hettype()` cannot be specified if choice sets are homogeneous.

`krobb(#)` allows the user to obtain Krinsky-Robb standard errors or confidence intervals using # draws.

`krse` specifies that standard errors are computed. The default is confidence intervals.

`krlevel(#)` specifies the confidence level. The default is 95%.

`krburn(#)` specifies the number of initial sequence elements to drop when creating the Halton sequences for the Krinsky-Robb procedure.

`kruserdraws` allows the user to provide draws for the Krinsky-Robb parametric bootstrap in a Mata matrix `mixl_uswerdraws`. The matrix must have the number of rows equal to the number coefficients in the model and the number of columns equal to the number of choice occasions times the number of repititions.[11] If `userdraws` is not specified, `mixlelast` will use Halton draws if the number of coefficients is less or eqal 10 and standard normal draws if it exceeds 10.[12]

`quietly` suppresses the output.

8 Examples

8.1 Homogeneous choice sets

To illustrate `mixlelast`, we use an artificial data set of 100 individuals making 10 repeated purchasing decisions. Each decision maker chooses from 5 different products.

The following three variables enter the choice model:

- product price
- product quality
- brand dummy

Following the notation of Hole (2007), `gid` identifies the alternatives in a choice occasion and `pid` identifies the choice occasion faced by a decision maker. The dependent variable `y` is 1 for the

[11] Note that here the number of coefficients is total number of coefficients estimated in `mixlogit`, i.e. the variable(s) specified in the `random()` option count twice as the mean and the SD are separate coefficients.

[12] Halton sequences are prefered with lower dimensions, however with dimension above 10, pseudo-random numbers outperform Halton draws, see Drukker and Gates (2006). More advanced procedures, such as sqrambled Halton sequences (Kolenikov 2012), can be used via the `kruserdraws` option.

chosen alternative and 0 otherwise.

The first 10 observations are listed below.

```
. use example_hom.dta
. list in 1/10,sepby(gid)

     +-------------------------------------------------------+
     |    alt   y      price   quality     brand   gid   pid |
     |-------------------------------------------------------|
  1. | Prod 1   0   1.308418         3   Brand A     1     1 |
  2. | Prod 2   0   2.357423         5   Brand A     1     1 |
  3. | Prod 3   0   3.380476         8   Brand A     1     1 |
  4. | Prod 4   1   4.414344        11   Brand B     1     1 |
  5. | Prod 5   0   8.617756        22   Brand B     1     1 |
     |-------------------------------------------------------|
  6. | Prod 1   0   .8786653         3   Brand A     2     1 |
  7. | Prod 2   0    1.99105         5   Brand A     2     1 |
  8. | Prod 3   0   3.536737         8   Brand A     2     1 |
  9. | Prod 4   0   4.295895        11   Brand B     2     1 |
 10. | Prod 5   1   8.679978        22   Brand B     2     1 |
     +-------------------------------------------------------+
```

We start by fitting a mixed logit model. The coefficient on price is assumed to be normally distributed, the coefficients on quality and the brand dummy are fixed.

```
. mixlogit y brand quality, rand(price) group(gid) id(pid)
Iteration 0:   log likelihood = -1386.9157
  (output omitted)
Iteration 4:   log likelihood = -1218.6785
Mixed logit model                               Number of obs     =      5,000
                                                LR chi2(1)        =     635.40
Log likelihood = -1218.6785                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Mean         |
       brand |   .8389558   .1664776     5.04   0.000     .5126657    1.165246
     quality |   .7326987   .0890225     8.23   0.000     .5582178    .9071797
       price |  -1.766318   .2272046    -7.77   0.000     -2.21163   -1.321005
-------------+----------------------------------------------------------------
SD           |
       price |   .5507636   .0545198    10.10   0.000     .4439067    .6576205
------------------------------------------------------------------------------
The sign of the estimated standard deviations is irrelevant: interpret them as
being positive
```

Both `price` and `quality` have the expected sign, i.e. the utility decreases in price and increases in quality. The SD coefficient is significantly different from zero, showing that there is preference heterogeneity. The positive coefficient on the dummy variable indicates that c.p. products of brand A (which is coded 1) are preferred.

We now compute unweighted sample point price elasticities.

```
. mixlelast, alternatives(alt)
Mixed logit sample elasticities
                  |  Prod 1      Prod 2      Prod 3      Prod 4      Prod 5
------------------+-----------------------------------------------------------
Prod 1            |
          Mean    | -1.583582     1.20702    1.234225    .5806123    1.160138
            SD    |  .5321263      .35009    .4233442    .2133211    .2015333
------------------+-----------------------------------------------------------
Prod 2            |
          Mean    |  .8620516   -2.993063    1.256074    .6522315    1.689042
            SD    |  .1310344    .6221173    .3964381     .226443    .2854372
------------------+-----------------------------------------------------------
Prod 3            |
          Mean    |  .6431518    .9099626   -4.794224    .7198147    2.664615
            SD    |  .1075382    .2245667    .6208259    .2306577    .4073909
------------------+-----------------------------------------------------------
Prod 4            |
          Mean    |  .4741855    .7339106    1.113647    -6.85574    3.730741
            SD    |  .0869767    .1709876    .2885865     .470882    .4400616
------------------+-----------------------------------------------------------
Prod 5            |
          Mean    |  .1097656    .2170911    .4650955    .4201686   -3.954723
            SD    |  .0208497    .0434097    .0906856    .1006895    .1834887
Calculated for a marginal change in price
```

The output table above displays the mean direct and cross elasticities and the respective standard deviations. The elements on the diagonal are the direct elastictities which are, as expected, all negative. Given a one percent increase in price, the decrease in the choice probability ranges from 1.584 to 6.856 percent. The off-diagonal elements are the cross elasticities, where each entry describes the percentage change in the row alternative given a one percent change in an attribute of the column alternative. E.g. the choice probability of Prod 1 increaes by 1.207 percent when the price of Prod 2 rises by 1 percent. All cross-price elasticities are positive indicating that the demand for rival alternatives increases as the price for one specific rival alternative rises.

Alternatively, we could compute probability weighted sample averages. This time, we make use of the `nosd` option which produces sample averages without standard deviations.

```
. mixlelast, alternatives(alt) weighted nosd
Mixed logit probability weighted sample elasticities
                  |  Prod 1      Prod 2      Prod 3      Prod 4      Prod 5
------------------+-----------------------------------------------------------
         Prod 1   | -1.437372    1.129342    1.178512    .5706147    1.147615
         Prod 2   |  .8389906   -2.795044    1.216929    .6372329     1.64468
         Prod 3   |  .6416162    .8983047   -4.591341    .7086891    2.560194
         Prod 4   |  .4834713    .7372868    1.113011   -6.706311    3.620954
         Prod 5   |  .1108344    .2183731    .4632295    .4174095   -3.941572
Calculated for a marginal change in price
```

Instead of calculating price elasticities, we could also consider a change in the quality variable. As price was the first (and only) variable set to be random in `mixlogit`, we did not have to specify it in `mixlelast`. To obtain elasticities w.r.t. quality, we need to use the `elastvar` option.

Quality is an integer variable and hence looking at a marginal change does not make very much sense. `mixlelast` will still produce results, but detects that quality is an integer variable and issues a warning. More meaningful, however, is to look at a discrete change in the quality variable, let's say by one.

```
. mixlelast, alternatives(alt) for(quality) absolutechange(1)
Mixed logit sample elasticities
             |    Prod 1     Prod 2     Prod 3     Prod 4     Prod 5
-------------+-------------------------------------------------------
Prod 1       |
        Mean |  1.388182  -1.396422  -1.524345  -.7976167  -1.520772
          SD |  .2607879   .4197096   .5221947   .2983788   .2449431
-------------+-------------------------------------------------------
Prod 2       |
        Mean | -1.005672   2.624413  -1.615071   -.929526  -2.236142
          SD |  .2040373    .388189    .513181   .3318357   .3566263
-------------+-------------------------------------------------------
Prod 3       |
        Mean | -.8210379   -1.18665   4.448009  -1.079768   -3.60011
          SD |  .1746568   .3130168    .459669   .3588448    .523858
-------------+-------------------------------------------------------
Prod 4       |
        Mean | -.6550803  -1.018255  -1.592809   7.037799  -5.179757
          SD |  .1483684   .2581882   .4276922   .3251452   .5807602
-------------+-------------------------------------------------------
Prod 5       |
        Mean | -.1922135   -.367616  -.7905778  -.7807326   4.480929
          SD |  .0462874   .0848141   .1714049   .2014827   .2330713
Calculated for an absolute change of 1 in quality
```

Contrary to price elasticities, we find positive direct and negative cross elasticities for quality. This is due to the positive sign on the quality coefficient and makes sense intuitively as an increase in quality should make a product more attractive.

By the same token, we could also look at the change in the brand variable, i.e. evaluate the counterfactual situation where a product is sold under the rival's brand. In this case, we would use the `dummy` option to simulate hypothetical changes of the dummy variable *brand* from zero to one and vice versa.

To obtain marginal effects instead of elasticities, we use the `marginaleffect` option.

```
. mixlelast, alternatives(alt) marginal
Mixed logit sample marginal effects
             |    Prod 1     Prod 2     Prod 3     Prod 4     Prod 5
-------------+-------------------------------------------------------
Prod 1       |
        Mean | -.2878743   .1365916   .0887652   .0314141   .0311034
          SD |  .0313022   .0341118   .0318789   .0144679   .0097729
-------------+-------------------------------------------------------
Prod 2       |
        Mean |  .1365916  -.2766283   .0749002   .0286624   .0364741
          SD |  .0341118   .0560073   .0303258   .0131542   .0110592
-------------+-------------------------------------------------------
Prod 3       |
        Mean |  .0887652   .0749002  -.2383686   .0268862   .0478171
          SD |  .0318789   .0303258   .0612841   .0126498   .0115994
-------------+-------------------------------------------------------
Prod 4       |
        Mean |  .0314141   .0286624   .0268862  -.1184339   .0314712
          SD |  .0144679   .0131542   .0126498   .0382954   .0084627
-------------+-------------------------------------------------------
Prod 5       |
        Mean |  .0311034   .0364741   .0478171   .0314712  -.1468658
          SD |  .0097729   .0110592   .0115994   .0084627   .0136522
Calculated for a marginal change in price
```

Here, the columns (and rows) add up to zero, as the decrease in the choice probability of one alternative is compensated by an increase in the probability of choosing the rival alternatives.

Finally, we look at a very small discrete change of 0.001 percent and recompute the marginal effects from above using the arc method.

```
. mixlelast, alternatives(alt) marginal percentchange(0.001)
Mixed logit sample incremental effects
                   |    Prod 1     Prod 2     Prod 3     Prod 4     Prod 5
-------------------+-------------------------------------------------------
Prod 1             |
              Mean | -.2878735   .1365904   .0887634    .031413   .0311024
                SD |  .0313027   .0341119   .0318786   .0144675   .0097726
-------------------+-------------------------------------------------------
Prod 2             |
              Mean |  .1365913  -.2766259   .0748988   .0286615   .0364732
                SD |  .0341118   .0560079   .0303255   .0131538   .0110589
-------------------+-------------------------------------------------------
Prod 3             |
              Mean |  .0887649   .0748996  -.2383644   .0268853   .0478164
                SD |  .0318789   .0303257   .0612842   .0126495   .0115992
-------------------+-------------------------------------------------------
Prod 4             |
              Mean |   .031414   .0286622   .0268857  -.1184302   .0314711
                SD |  .0144679   .0131541   .0126497   .0382946   .0084626
-------------------+-------------------------------------------------------
Prod 5             |
              Mean |  .0311032   .0364737   .0478164   .0314704  -.1468632
                SD |  .0097729   .0110592   .0115994   .0084626   .0136522
Calculated for a change of .001 per cent in price
```

As expected, the results for sample averages and their standard deviations are practically identical. Nonetheless, the researcher is advised to use the arc method only when necessary: the computation is somewhat more complicated and hence more time consuming. The difference in terms of computation time is both increasing in the number of alternatives and the number of choice occasions.

8.2 Heterogeneous choice sets

In the second data set, decision makers live in one of three locations, which differ according to the set of products available. Indvididuals 1 to 50 live in Location 1 where all five product of both brands A and B are available. In Location 2, where individuals 51 to 75 live, only products of brand A are available, i.e. products 1, 2 and 3. Finally, only Products 4 and 5 of Brand B are in the choice sets of decision makers 75 to 100 living in Location 3.

A representative choice set for each location is displayed below:

```
. use example_het.dta
. list if gid == 1 | gid == 501 | gid == 751, sepby(gid)

      +----------------------------------------------------------------------+
      |    alt   y      price   quality     brand   gid   pid   location     |
      |----------------------------------------------------------------------|
   1. | Prod 1   0   1.308418         3   Brand A     1     1          1     |
   2. | Prod 2   0   2.357423         5   Brand A     1     1          1     |
   3. | Prod 3   0   3.380476         8   Brand A     1     1          1     |
   4. | Prod 4   1   4.414344        11   Brand B     1     1          1     |
   5. | Prod 5   0   8.617756        22   Brand B     1     1          1     |
      |----------------------------------------------------------------------|
2501. | Prod 1   0   1.356619         3   Brand A   501    51          2     |
2502. | Prod 2   1   2.180957         5   Brand A   501    51          2     |
2503. | Prod 3   0    3.14988         8   Brand A   501    51          2     |
      |----------------------------------------------------------------------|
3251. | Prod 4   1   4.398738        11   Brand B   751    76          3     |
3252. | Prod 5   0   8.765503        22   Brand B   751    76          3     |
      +----------------------------------------------------------------------+
```

As before, we estimate the mixed logit model.

```
. mixlogit y brand quality, rand(price) group(gid) id(pid)
Iteration 0:   log likelihood = -1094.6025
  (output omitted)
Iteration 5:   log likelihood = -930.11945
Mixed logit model                               Number of obs     =      3,750
                                                LR chi2(1)        =     559.92
Log likelihood = -930.11945                     Prob > chi2       =     0.0000

------------------------------------------------------------------------------
           y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Mean         |
       brand |   .9619626   .2485758     3.87   0.000      .474763    1.449162
     quality |   .8084889   .1005983     8.04   0.000     .6113199    1.005658
       price |  -1.926808   .2566438    -7.51   0.000    -2.429821   -1.423795
-------------+----------------------------------------------------------------
SD           |
       price |   .6458248   .0680799     9.49   0.000     .5123906     .779259
------------------------------------------------------------------------------
The sign of the estimated standard deviations is irrelevant: interpret them as
being positive
```

First, we compute the unweighted sample elasticities of type I, i.e. over the entire sample of 100 decision makers.

```
. mixlelast, alternatives(alt) hettype(I) nosd
Mixed logit sample elasticities (type I)
             |    Prod 1     Prod 2     Prod 3     Prod 4     Prod 5
-------------+-------------------------------------------------------
      Prod 1 | -1.249679   1.027456   1.100893   .2595871   .5574325
      Prod 2 |  .7087478  -2.334087   1.184107   .3049821   .8680116
      Prod 3 |  .4862719   .7379635  -3.355797   .3496448   1.460553
      Prod 4 |  .2535825   .3996399   .6194909  -4.395615   3.301949
      Prod 5 |  .0487804   .1000686   .2244125   .6515905   -2.76727
Calculated for a marginal change in price
```

Now we restrict our attention to those individuals who have the row product in their choice set (type IIa).

```
. mixlelast, alternatives(alt) hettype(IIa) nosd
Mixed logit sample elasticities (type IIa)
```

	Prod 1	Prod 2	Prod 3	Prod 4	Prod 5
Prod 1	-1.666239	1.369941	1.467857	.3461161	.7432433
Prod 2	.9449971	-3.112116	1.57881	.4066428	1.157349
Prod 3	.6483625	.9839513	-4.474397	.4661931	1.947404
Prod 4	.33811	.5328532	.8259879	-5.86082	4.402598
Prod 5	.0650405	.1334248	.2992166	.8687874	-3.689693

```
Calculated for a marginal change in price
```

Comparing both tables shows that direct and cross elasticities have increased in absolute values. This result is not surprising. For type I, we consider all individuals including some whose individual elasticities are zero. For IIa, we ignore all individuals who could not have chosen the row product. That is for products 1,2 and 3 we now look at decision makers in Location 1 and Location 2 only. Likewise, only decision makers in Location 1 and Location 3 are considered for products 4 and 5.

This means that we no longer consider individuals with zero individual direct elasticities, which is why direct sample elasticities must rise in absolute value. For cross elasticities, we also reduce the number of decision makers with zero individual elasticities, namely those who could not choose the row product. Note, however, that we still consider those instances of zero individual cross elasticities, where the row product is available but the column product is not.

This leads us to type IIb where we only consider those individuals who are actually affected by an attribute change, i.e. both the row and the column alternative are present in their choice set.

```
. mixlelast, alternatives(alt) hettype(IIb) nosd
Mixed logit sample elasticities (type IIb)
```

	Prod 1	Prod 2	Prod 3	Prod 4	Prod 5
Prod 1	-1.666239	1.369941	1.467857	.5191742	1.114865
Prod 2	.9449971	-3.112116	1.57881	.6099641	1.736023
Prod 3	.6483625	.9839513	-4.474397	.6992897	2.921106
Prod 4	.507165	.7992798	1.238982	-5.86082	4.402598
Prod 5	.0975608	.2001371	.4488249	.8687874	-3.689693

```
Calculated for a marginal change in price
```

Comparing the results for type IIa and type IIb, we note three things. First, direct elasticities are identical. This is because the subset of individual direct effects taken into consideration does not differ between type IIa and type IIb. Second, cross elasticities between alternatives of the same brand also remain unchanged. Recalling that the heterogeneity was on brand level, there are no cases in which one product of brand A is available and another is not. Hence, regarding substitution patterns between alternatives of the same brand, there is no difference between type IIa and type IIb either. Third, for cross effects across brands, only those decision makers who have both products available remain. All individuals with zero cross effects are now excluded which is why the sample effects rise in absolute value.

8.3 Standard errors and confidence intervals

Finally, we want to compute confidence intervals and standard errors using the Krinsky-Robb parametric bootstrap. For demonstration, we use the same data set as in the first example, see subsection 8.1.

We now compute standard errors with 100 replications using the `krse` option.

```
. mixlelast, alternatives(alt) kr(100) krse
Mixed logit sample elasticities
```

		Prod 1	Prod 2	Prod 3	Prod 4	Prod 5
Prod 1						
	Mean	-1.588732	1.209058	1.235619	.5851914	1.168598
	SE	.1741421	.1293064	.14276	.0947962	.2206916
Prod 2						
	Mean	.8656952	-3.006101	1.258633	.657472	1.695867
	SE	.1059331	.3317349	.1474431	.1033878	.283632
Prod 3						
	Mean	.6456213	.9128968	-4.814247	.7260494	2.669903
	SE	.076321	.1053874	.5710928	.1146381	.4180427
Prod 4						
	Mean	.4762443	.7370268	1.118481	-6.881942	3.73549
	SE	.0566474	.0891121	.1443512	.8692364	.5892882
Prod 5						
	Mean	.1115721	.2204908	.4716262	.4269033	-4.017563
	SE	.0207407	.0412111	.0887768	.0860485	.7422291

```
Calculated for a marginal change in price
Means and standard errors by Krinsky-Robb parametric bootstrap with 100 repetitions
```

Alternatively, we could compute confidence intervals. The default is 95% confidence intervals. To set the confidence level to, let's say, 99%, we make use of the `krlevel` option.

```
. mixlelast, alternatives(alt) kr(100) krlevel(99)
Mixed logit sample elasticities
                 |    Prod 1      Prod 2      Prod 3      Prod 4      Prod 5
-----------------+-----------------------------------------------------------
Prod 1           |
            Mean | -1.588732    1.209058    1.235619    .5851914    1.168598
        CI_lower | -2.036889     .885817    .9136785     .390892    .7357463
        CI_upper | -1.162032    1.540968    1.617532    .8230219    1.778616
-----------------+-----------------------------------------------------------
Prod 2           |
            Mean |  .8656952   -3.006101    1.258633     .657472    1.695867
        CI_lower |   .594959   -3.751526    .9098775    .4391683    1.131964
        CI_upper |  1.192836   -2.167427    1.635249    .9178194    2.445996
-----------------+-----------------------------------------------------------
Prod 3           |
            Mean |  .6456213    .9128968   -4.814247    .7260494    2.669903
        CI_lower |  .4488157    .6570522   -6.078871    .4858516    1.739164
        CI_upper |  .8719096    1.181786   -3.407677    1.006442    3.725657
-----------------+-----------------------------------------------------------
Prod 4           |
            Mean |  .4762443    .7370268    1.118481   -6.881942     3.73549
        CI_lower |  .3345948    .5277493    .7775908    -8.96048    2.333426
        CI_upper |  .6314298    .9583172    1.492383   -4.728254    5.122632
-----------------+-----------------------------------------------------------
Prod 5           |
            Mean |  .1115721    .2204908    .4716262    .4269033   -4.017563
        CI_lower |  .0712387     .141619    .2904997    .2830188   -6.543079
        CI_upper |  .1802272    .3584905    .7616104    .6880237   -2.611262
Calculated for a marginal change in price
Means and 99% confidence intervals by Krinsky-Robb parametric bootstrap with 100 repetitions
```

As we can see, none of the confidence intervals contains zero and hence all direct- and cross-elasticities are significantly different from zero at the 1% level. However, to obtain robust results, we should use a much higher number of draws.

9 Saved results

`mixlogit` stores the following results to r():

Scalars

`r(N)`	number of observations	`r(burn)`	initial elements to drop when creating Halton sequences
`r(N_group)`	number of choice occasions	`r(nrep)`	number of Halton draws
`r(N_id)`	number of decision makers	`r(krobb)`	number of Krinsky-Robb repetitions
`r(N_alt)`	number of alternatives	`r(krlevel)`	Krinsky-Robb confidence level
`r(het)`	1 if heterogeneous choice sets, 0 otherwise	`r(krse)`	1 if Krinsky-Robb standard erros, 0 if confidence intervals
`r(marginal)`	1 if marginal effects, 0 otherwise	`r(krburn)`	initial elements to drop when in Halton seqence for `krobb`
`r(weighted)`	1 if probability weighted effects, 0 otherwise	`r(kruser)`	1 if user-provided random numbers, 0 otherwise
`r(nosd)`	1 if no standard deviations, 0 otherwise		

Macros

`r(cmd)`	`mixlelast`	`r(title)`	title in output table
`r(subtitle)`	subtitle in output table	`r(altid)`	name of altid() variable
`r(subtitle2)`	subtitle of output for `krobb`	`r(hettype)`	type of aggregation for heterogeneous choice sets
`r(elastvar)`	name of elastvar() variable	`r(group)`	name of group() variable
`r(id)`	name of id() variable	`r(method)`	type of calculation method

Matrices

`r(mean)`	sample elasticities/marginal effects	`r(sd)`	standard deviations
`r(KRSE)`	Krinsky-Robb standard errors	`r(KR_lower)`	Krinsky-Robb lower confidence bounds
`r(KR_upper)`	Krinsky-Robb upper confidence bounds		

References

Cameron, A. C. and Trivedi, P. K. (2010), 'Microeconometrics using stata'.

Drukker, D. M. and Gates, R. (2006), 'Generating halton sequences using mata', *Stata Journal* **6**(2), 214–228.

Greene, W. H. (2008), *Econometric Analysis*, Prentice Hall.

Hensher, D. A., Rose, J. M. and Greene, W. H. (2015), *Applied Choice Analysis*, Cambridge University Press.

Hole, A. R. (2007), 'Estimating mixed logit models using maximum simulated likelihood', *Stata Journal* **7**(3), 388–401.

Kolenikov, S. (2012), 'Scrambled halton sequences in mata', *Stata Journal* **12**(1), 29.

Krinsky, I. and Robb, A. L. (1986), 'On approximating the statistical properties of elasticities', *The Review of Economics and Statistics* pp. 715–719.

Krinsky, I. and Robb, A. L. (1990), 'On approximating the statistical properties of elasticities: a correction', *The Review of Economics and Statistics* **72**(1), 189–90.

Louviere, J. J., Hensher, D. A. and Swait, J. (2000), *Stated Choice Methods: Analysis and Applications in Marketing, Transportation and Environmental Evaluation*, Cambridge University Press.

McFadden, D. (1973), Conditional logit analysis of qualitative choice behavior, *in* 'Frontiers in Economics', Academic Press.

Revelt, D. and Train, K. (1998), 'Mixed logit with repeated choices: Households' choices of appliance efficiency level', *The Review of Economics and Statistics* **80**(4), pp. 647–657.

Train, K. (2009), *Discrete choice methods with simulation*, Cambridge University Press.