Specification tests for the multinomial logit model revisited: The role of alternative-specific constants

Revision: January 29, 2018

Jan Rouwendal[1]

[1] Department of Spatial Economics, Vrije Universiteit; Tinbergen Institute, Netherlands.

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at http://www.tinbergen.nl

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

Specification tests for the multinomial logit model revisited:
The role of alternative-specific constants

Jan Rouwendal[1,2]

Key words multinomial logit, Hausman-McFadden specification test, independence of irrelevant alternatives, nested logit, GEV models, mixed logit
JEL codes D1 D4

Abstract
This paper shows that Hausman-McFadden and related tests for the multinomial logit model are not informative about failure of IIA, but only about specification error in the deterministic part of the utilities attached to choice alternatives. It is shown that the multinomial logit cannot be distinguished from any more general additive random utility model when a full set of alternative specific constants is present. Estimation of such more general models only addresses specification error in the deterministic parts of the utilities and is not informative about the true specification of the random terms of the utilities.

[1] Department of Spatial Economics, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, Netherlands.
[2] Tinbergen Institute, Gustav Mahlerplein 117, 1082 MS Amsterdam, Netherlands.

1 Introduction

A vast literature in industrial organization, marketing, transportation and other fields estimates discrete choice models that generalize the multinomial logit model (MNL) by allowing for correlation between the random parts of the utilities. Examples include Small's (1987) ordered alternatives model, Goldberg's (1995) nested logit model with multiple layers, Bresnahan, Stern, & Trajtenberg's (1997) Product Differentiation GEV model with overlapping nests. The main reason for using these models is the restrictive independence of irrelevant alternatives (IIA) property of MNL. Generalizations like the nested, cross-nested or mixed logit models[1] all avoid this property to some extend and are therefore regarded as being superior in that they reflect reality better than the MNL does. However, in this paper I show that these models are only superior to MNL in the sense that they give a better approximation to an MNL that has a full set of alternative-specific constants for each type of actor. The difference between the chosen specification of the random parts of the utilities and that of the MNL is only statistically significant because it helps to address omitted variable bias in the deterministic parts of the utilities, but it is not informative about the true specification of the random parts.

This conclusion is reached through examination of the specification tests for the MNL that were developed by Hausman & McFadden (1984) with the intention to examine the validity of IIA. I show that these tests lose all their power if an MNL with a full set of alternative-specific constants is estimated. Since these constants absorb all elements of the deterministic parts of the utilities, failure to pass the test indicates specification error in the deterministic parts of the utilities. However, the tests tell us nothing about the specification of the random parts of the utilities.

The likelihood function of a nested logit model becomes flat in the parameters that embody a deviation from MNL once the alternative-specific constants are introduced. This conclusion generalizes to any generalized extreme value (GEV) model and to mixed logit models. If different types of actors can be distinguished, these conclusions continue to hold if alternative specific constants are used for each type of actor. In other words, once we have eliminated specification error in the deterministic parts of the utilities by absorbing it in the alternative-specific constants, there is nothing left in the data to inform us about the true specification of the random parts.

The implication is that generalizations of MNL that are superior in terms of the likelihood only appear to be so because they offer a closer approximation to the saturated MNL with a full set of alternative specific constants. However, such models do not tell us anything about the true specification of the random parts of the utilities that is in the data. This means that a large set of studies that have claimed to find violations of IIA and superior substitution patterns due to correlations between random parts of the utilities for various alternatives have not done so on solid empirical grounds.

The paper unfolds as follows. In the next section the background of the analysis is sketched. In section 3 the introduction of alternative-specific constants in MNL will be discussed. Sections 4 and 5 consider the tests proposed by Hausman & McFadden (1984) when these constants are introduced. Since they absorb all elements of the deterministic utilities, applying the tests in this context implies restricting them to the random, or idiosyncratic, parts of the utilities that determine the validity or violation of IIA. Section 6 discusses other possible specification tests of MNL. Section 7 considers heterogeneity of the actors. Section 8 concludes.

---

[1] See McFadden (1978), Abbe, Bierlaire, & Toledo (2007), McFadden & Train (2000), respectively.

1

2 Background

The MNL is still the most popular discrete choice model in economic research. The constituting property of this model is IIA (Luce, 1959), which is also generally regarded as its main weakness. The best known aspect of IIA is the 'red bus – blue bus' problem, according to which the introduction of a duplicate of an existing choice alternative – that does not essentially change the choice menu – can have substantial impact on all choice probabilities (Debreu, 1960). Perhaps even more important is the fact that, with identical actors, a change in attractiveness of one alternative implies substitution behavior towards or from the other alternatives that is determined in a mechanical way by the choice probabilities.

McFadden (1974) famously developed an economic interpretation of the MNL that is based on the maximization of random utility. In this approach the utility attached to the alternatives is the sum of a deterministic part and a random part. The latter reflects idiosyncratic preferences unknown to the researcher. He showed that the logit model results if and only if these random parts are independent and identically distributed (iid) extreme value type I (EV I) distributed. The red bus – blue bus paradox arises from the fact that the deterministic parts of the utilities of the original alternative and its duplicate are identical, while the random parts are independent draws from the EV I distribution. The two types of buses would only really be duplicates if their random parts would also be identical for each consumer.

The restrictiveness of the IIA property – and therefore of the MNL – is ultimately an empirical question. In some situations it may be reasonable, in others not. This suggests the development of tests for its validity in actual situations. McFadden (1978) showed that the coefficients of the logit model can be recovered from estimating the model on choices that are restricted to a subset of the alternatives, provided that the model is correctly specified. This hints at a specification test of the MNL that is very suggestive of being a direct test of the IIA property. Hausman & McFadden (1984) further explored this issue in two directions. One is by elaborating the estimation on a subset of alternatives as a Hausman (1978) specification test.

McFadden (1978) generalized the MNL framework to situations in which the idiosyncratic utilities can be correlated. This generalized extreme value (GEV) framework encompasses a wide class of models in which the idiosyncratic utilities of subsets of alternatives can be correlated. The nested logit is the best known member of it and it has the MNL model as a special case that occurs when the parameters for the 'inclusive value' of the nests are equal to 1. Estimating a nested logit model can therefore be regarded as doing a specification test for the MNL and this is the second direction in which Hausman & McFadden (1984) explored the possibilities to test for IIA.

More recent work using MNL has emphasized the potential importance of unobserved (by the researcher) characteristics of the choice alternatives. Berry, Levinsohn, & Pakes (1995) (BLP) in particular have emphasized that neglecting such omitted variables is associated with an 'overfitting' problem. Moreover, explicitly taking into account their possible presence clarifies an important endogeneity problem that may severely bias the coefficient for the prices associated with the alternatives. To address these issues, they introduce alternative-specific constants (*asc*s) in the logit model that absorb the impact of all – observed as well as unobserved – characteristics for the average agent. In a second estimation stage, the *asc*s can be decomposed using linear regression methods. The error term includes the impact of the unobserved characteristics.

It may be conjectured that the neglect of unobserved characteristics of alternatives also has an impact on tests for IIA. The unobserved characteristics are included in models following BLP as a separate error term that takes on the same value for all agents. No distributional assumptions are made for this term. It is conceptually clearly distinguished from the idiosyncratic

2

error terms that are considered to be independent draws from the extreme value type I distribution for all agents. However, ignoring the possible presence of unobserved characteristics means that the unobserved characteristics are not separated from the random parts of the utilities. This will result in a composite error term that may be correlated among the agents. Moreover, if two alternatives have similar unobserved characteristics the composite error terms will be correlated across the alternatives even if the idiosyncratic utilities satisfy the assumptions of MNL. This suggests that the specification tests of Hausman & McFadden (1984) concern the *joint* hypothesis of IIA and the absence of unobserved characteristics unless one properly controls for the latter. In this paper I will show that if one does so, the tests do not have *any* power for detecting a possible violation of IIA. To make this point as clearly as possible, I consider a population of agents whose deterministic parts of the utilities attached to a given choice alternative are identical. That is, the agents only differ in the random parts of their preferences. These idiosyncratic utilities may or may not obey the assumptions of the MNL and we will consider whether estimating the MNL on subsets or estimating a nested logit model will, in this situation, give us information about the validity of the assumption that IIA is present.

The recent literature has downplayed the role of IIA and emphasized that, provided enough heterogeneity of choice behavior is permitted by the model, the empirical substitution patterns are determined by the estimated coefficients, rather than by the model specification. Be that as it may, there is another reason why testing for IIA is important. McFadden (1984) argues that the MNL functional form is not restrictive if one allows the 'utilities' of the alternatives to depend on the characteristics of all alternatives: The universal or 'mother' logit can approach any discrete choice model as close as desired. It is thus the exclusion of the characteristics of the other alternatives that gives the IIA property its bite. If this property is absent, we cannot interpret the estimated *asc*s as reflecting the values of the deterministic parts of the utilities of the associated alternatives. Instead, they can be functions of all the deterministic utilities and hence of the characteristics of all alternatives. The exclusion of the characteristics of all other alternatives is the cornerstone of BLP's identification strategy as it motivates the choice of their instruments for the price. The validity of this strategy thus depends on that of the IIA property.

3 The setting

We consider a population of actors $h = 1 \dots H$ with additive random utility functions that have to choose among N alternatives indexed $n = 1 \dots N$. The utility attached to alternative n is:

$$u_n^h = v_n + \varepsilon_n^h, \ n = 1 \dots N; h = 1 \dots H \tag{1}$$

In this equation v_n is the deterministic part of the utility, which is equal for all actors, and ε_n^h is the random or idiosyncratic part that differs among actors. The logit model results from utility maximizing behavior if these random parts are iid EV I distributed. McFadden (1974) derives the choice probabilities as:

$$\pi_n = \frac{e^{v_n}}{\sum_{m=1}^N e^{v_m}} \tag{2}$$

That is, IIA results if this assumption is valid and not if the random parts of the utilities are distributed differently. Testing IIA thus means testing the validity of the assumption that the idiosyncratic parts of the utilities are iid EV I distributed.

The deterministic part of utility is often specified as a linear function of explanatory variables $X_1 \dots X_k \dots X_K$. At least since the publication of BLP it has been realized that such a specification may suffer from omitted variables. To take these into account, one may write:

$$v_n = \sum_{k=1}^K \beta_k X_{kn} + \xi_n, \ n = 1 \dots N. \tag{3}$$

where ξ_n denotes the impact of characteristics of the choice alternative that are not observed by the researcher. Since the ξ_n's are unobserved they cannot be taken into account directly. However, on the basis of (2), one can estimate the v_n's, that absorb the ξ_n's, as *asc*s that are determined up to an additive term by the observed choice probabilities. The coefficients β in (3) may be recovered in a second estimation step, using the estimated *asc*s as dependent variables. In this second stage the ξ_n's function as error terms.

The earlier, pre-BLP, literature does not take into account the possibility of unobserved characteristics in the deterministic part of utility and may therefore be interpreted as referring to a model in which the impact of such characteristics is included in the error term:

$$u_n^h = \underbrace{\left(\sum_{k=1}^{K} \beta_k X_{kn}\right)}_{v'_n} + \underbrace{(\xi_n + \varepsilon_n^h)}_{\varepsilon'^h_n} \tag{4}$$

The presence of a nonzero ξ_n introduces correlation between the error terms ε'^h_n. Since ξ_n is equal for all individuals the ε'^h_n's become correlated among the individuals for a given n. If the unobserved variables are correlated between the alternatives (for instance, because the same characteristic is present in a number of choice alternatives), there will be correlation between the ε'^h_n's for different n's as well. Introduction of *asc*s as proposed by BLP thus allows one to make a distinction between the impact of unobserved characteristics that are included in the ξ_n's and that of the idiosyncratic terms ε_n^h. Without *asc*s the specification tests refer to the hypothesis that $\varepsilon'^h_n = (\xi_n + \varepsilon_n^h), n = 1 \dots N, h = 1 \dots H$ is iid EV I distributed, which is equal to the joint hypothesis that $\xi_n = 0$ and ε_n^h is iid EV I. With *asc*s, the test refers only to the latter hypothesis. Since the IIA property is present when the latter hypothesis holds true, and does not depend on the validity of the first, testing for this property when including alternative-specific constants is the preferred procedure. In the next sections we will therefore consider how the IIA-tests perform in this context.

4 Estimating MNL on subsets

Hausman & McFadden (1984) consider estimation of a logit model with an arbitrary number N of choice alternatives. They work with logit model (2) in which the deterministic parts of the utilities are linear functions of K explanatory variables $X_1 \dots X_k \dots X_K$:

$$v_n = \sum_{k=1}^{K} \beta_k X_{kn}, \quad n = 1 \dots N. \tag{5}$$

Note the absence of the ξ_n's in this equation. McFadden (1978) shows that consistent estimates of the parameters β can be obtained from a fixed or random sample of the full choice set. It is somewhat intuitive that this property is related to IIA, which requires that the ratio of two choice probabilities does not depend on the presence of other alternatives. Following this line of thought, McFadden, Train, & Tye (1977) proposed a diagnostic test for IIA based on this property. Hausman & McFadden (1984) developed this methodology further by linking it to Hausman's (1978) specification test.[2,3]

The test works by comparing the estimates of β that result from using all observations with the estimates that result when a subset of observations – only those referring to a subset of the choice alternatives - is used. If they are not significantly different, the model passes the test and it is concluded that IIA holds.[4]

---

[2] See also Small & Hsiao (1985).
[3] Very recently Hahn, Hausman, & Lustig (2017) provided an extension of the test to mixed logit models.
[4] It may be the case that some parameters are relevant only for alternatives that are not included in the second estimation. The test refers only to the subset of parameters that are included in both estimations.

The intuitive argument discussed above notwithstanding, it is surprising that one could be able to test the presence of IIA by removing some of the choice alternatives from data about choice behavior that refers to a situation in which they are all available.[5] This makes one wonder how exactly the test works.

To see that the test does indeed not only refer to IIA, consider the case in which the deterministic part of the utilities attached to the alternatives is not correctly specified. More specifically, assume that there is an omitted variable that only affects the utilities of a subset of the alternatives. If this subset coincides with the subset that is chosen for testing IIA, and the omitted variable is correlated with the variables that are included in the model, it is clearly possible that the Hausman-McFadden test rejects IIA, while this property may in fact be valid in the correctly specified model. In this example the Hausman-McFadden test rightly rejects the chosen specification, but because of the misspecification of the deterministic part of the utility function, not because of a failure of IIA.

BLP have emphasized that omitted variables in the logit model will give rise to an "embarrassing" over-fitting problem: if the model were correctly specified differences between the observed choice probabilities and those predicted by the model can only be due to sampling error, while in reality the differences are often much larger than can be explained by this source (p. 850). Estimation of *asc*s solves this issue because it makes the predicted choice frequencies identically equal to the observed ones.

However, with a homogeneous population the estimates of the *asc*s are the same, whether they are estimated on the whole set of observations or on a subsample with choices referring to a subset of alternatives only. The reason is the one-to-one correspondence between the estimated *asc*s and the choice frequencies that is present in this case.

To see this, assume we have F observations of choices among alternatives $1 \ldots N$. The model we estimate is:

$$\pi_n = \frac{e^{asc_n}}{\sum_{m=1}^{N} e^{asc_m}}, \quad n = 1 \ldots N. \tag{6}$$

Let *A* denote the subset of alternatives used in the estimation. The loglikelihood is:

$$L_A = \sum_{n \in A} f_n (asc_n - \ln(\sum_{m \in A} e^{asc_m})) \tag{7}$$

where f_n is the number of observations choosing alternative n. Maximization w.r.t. asc_n gives:

$$f_n = \frac{e^{asc_n}}{\sum_{m \in A} e^{asc_m}} F_A \tag{8}$$

with $F_A = \sum_{n \in A} f_n$.

For the maximum likelihood estimates of the asc's of two alternatives, n ad m, it must then be true that:

$$asc_n - asc_m = \ln f_n - \ln f_m. \tag{9}$$

Up to an additive constant, the *asc*s are thus determined by the choice frequencies and independent of the chosen subset A. The value of the constant is arbitrary. There are therefore no differences in the estimates of the *asc*s that can provide information about the validity of IIA, or any other aspect of the MNL.

The fact that this result refers to a population of homogeneous actors serves to emphasize that this Hausman-McFadden test is *not* informative about the (in)validity of IIA. McFadden et al. (1981) show that in a population with heterogeneous actors one cannot expect IIA to hold. Any test of this property should therefore focus on the behavior of actors that – according to the null hypothesis – have identical deterministic parts of their utilities.

---

[5] Allison (2012) relates that this bothered him.

When *asc*s are introduced and the population is homogeneous, the model is in fact saturated. The only information that is available is the choice of an alternative and this information is fully exploited by the *asc*s. In this framework there is nothing more to discover about the behavior of the agents.[6]

The significance of the result just reached can be further clarified by realizing that the inclusion of *asc*s controls for *all* factors that affect the deterministic utility in a homogeneous population. It is similar to the inclusion of 'fixed effects' in an OLS or panel data regression. What has been shown in this section is that, once omitted variable bias and specification errors have been excluded in this way the specification test loses all its power, although IIA can still be violated. Hence the first Hausman-McFadden test cannot be regarded as a proper test for IIA.

5 Estimating a nested logit model

McFadden (1978) presents the class of generalized extreme value (GEV) models in which the random parts of the utilities still have marginal distributions that are extreme value type I, but do not need to be independent. Relaxing the independence of the random parts implies that the IIA property is no longer present. The best known member of the GEV class, apart from the MNL, is the nested logit model. In this model each choice alternative is assigned to a group, or nest,[7] and the random parts of the utilities of alternatives belonging to the same group are allowed to be correlated. Conditional on the choice of a nest, choice behavior is still adequately described by an MNL and IIA holds. Similarly, the choice between nests can be described by another MNL. The parameters of this second MNL reflect the possible deviation from the conventional non-nested MNL. See, for instance, Train (2003) for further discussion of the model.

The MNL is a special case of the nested logit model in which the parameters of the inclusive values of the nests – which reflect the correlation between the random parts of the utilities of the alternatives included in the nest – are equal to 1. This suggests the second possibility for testing the IIA property that was also explored by Hausman & McFadden (1984): estimate a nested logit model and check if the parameters of the inclusive values differ significantly from 1. If so, IIA has to be rejected.

However, I will now show that in a homogeneous population where the *asc*s represent the deterministic parts of the utilities the likelihood function is flat in the coefficients for the inclusive values of the nests. Just as was the case for the test based on estimation on subsets of the choice alternatives, the test that uses the nested logit model has no power for detecting a violation of IIA in the situation where it has been isolated as the only remaining specification issue.

Some additional notation is necessary to formally introduce the nested logit. I use $k = 1 \ldots K$ to refer to the nests and $k(n)$ to indicate the nest to which alternative n belongs. The set of alternatives belonging to nest k is B_k. Finally, the parameters for the inclusive values are denoted as λ_k. The choice probabilities are:

$$\pi_n = \frac{e^{v_n/\lambda_{k(n)}} \left(\sum_{m \in B_{k(n)}} e^{v_m/\lambda_{k(n)}} \right)^{\lambda_{k(n)}-1}}{\sum_{k=1}^{K} \left(\sum_{m \in B_k} e^{v_m/\lambda_k} \right)^{\lambda_k}} \tag{10}$$

In Appendix A I formulate the likelihood function associated with the model and derive the first order conditions for maximizing it with respect to the *asc*s and the parameters for the inclusive

---

[6] It may be noted that this is the situation in which only aggregate data on choices (market shares) are known. Estimation of the BLP model is discussed in section 7.

[7] Groups are allowed to contain just one alternative.

values. I show that if the former are equal to zero – as is necessary for a maximum – the latter will automatically be also equal to zero. The likelihood function is thus flat in the parameters for the inclusive values, once the first order conditions for the *asc*s are satisfied. The implication of this result is that the nested logit model does not have *any* power for detecting IIA once the possibility of omitted variables in the specification of the deterministic parts of the utilities is properly taken into account.[8]

6 Other tests of IIA

6.1 Other GEV models
It is shown in Appendix B that the result derived for the nested logit generalizes to *any* GEV model. In all these models optimal choice of the *asc*s implies that the observed choice frequencies are equalized to the choice frequencies of the estimated model, and this then implies that the likelihood becomes flat with respect to the choice of the parameters of the GEV distribution.

6.2 Mixed logit models
The mixed logit model is sometimes considered to be the extension of the logit model that solves essentially all its limitations. In particular, McFadden and Train (2000) have shown that the model is able to approximate *any* discrete choice model. We will now consider the question whether the mixed logit model is able to distinguish the MNL from a model in which the actors have idiosyncratic utilities that are not iid EV I distributed. To do this, we assume that the utility attached to alternative n by actor h, u_n^h, equals:

$$u_n^h = asc_n + \vartheta_n^h + \varepsilon_n^h \tag{11}$$

where ε_n^h represents the standard MNL iid EV I error term and ϑ_n^h is another random variable. It is important to note that (11) overcomes the limitations of the iid EV I assumption as the random parts of the utilities are now equal to $(\vartheta_n^h + \varepsilon_n^h)$ which are unrestricted.

The common case is that ϑ_n^h is further specified as the product of a parameter ρ_n to be estimated and a random variable θ_n whose distribution is known, for instance a standard normal variate. The utility function then becomes:

$$u_n^h = asc_n + \rho_n \theta_n^h + \varepsilon_n^h \tag{12}$$

Estimation of this model can be regarded as doing a specification test for the MNL: if all ρ_n's are equal to zero, MNL is correct and IIA holds. If not, IIA has to be rejected. The deviations of the idiosyncratic utilities may be related to characteristics of the agents that are unobserved by the researcher, or that she is unwilling to use, for instance because she wants to get an idea of the total amount of heterogeneity in preferences present in the population before investigating it further. The choice probabilities can be written as:

$$Pr_n = \int_{\boldsymbol{\theta}} \frac{e^{asc_n + \rho_n \theta_n}}{\sum_{m=1}^{N} e^{asc_m + \rho_m \theta_m}} g(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

---

[8] There is a connection between the results of the present and previous sections. To see this consider the sequential estimation of a two-level nested logit model. The first step is to estimate logit models for each nest, which is identical to estimation of the logit model on the subsets of alternatives defined by the nest. In the previous section it was shown that this gives results that are equivalent to those of estimating the complete model. The 'logsums' of the nest are plugged into the logit model that concerns the choice of the nest. But then this upper level model is identical to the nonnested logit model if the logsum coefficients are all equal to one. We know that setting the coefficients at this value maximizes the likelihood..

$$= \int_{\boldsymbol{\theta}} \pi_n(\boldsymbol{\theta})g(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{13}$$

In this equation $\boldsymbol{\theta}$ denotes the $Nx1$ vector with elements θ_n, $g(\boldsymbol{\theta})$ the multidimensional density function of these random variables and $\pi_n(\boldsymbol{\theta})$ is shorthand notation for the logit expression in the first line. The loglikelihood ll_n if an observed choice for n is $\ln Pr_n$ and the likelihood of the whole set of observations is:

$$LL = \sum_n f_n \ln Pr_n \tag{14}$$

It is shown in Appendix C that LL is flat in the parameters $\rho_n, n = 1 \dots N$ as soon as the first-order conditions for the alternative-specific constants are satisfied. That is, the mixed logit model cannot distinguish between MNL and an additive random utility model in which idiosyncratic utilities differ from iid EV I. Put differently, estimating a mixed logit model of the type just described does not reveal any information about the presence of heterogeneity in the population that is not captured by the standard logit term.

Note that the random variables θ_n are allowed to be dependent. Moreover, the proof generalizes trivially to cases in which there is more than one random coefficient, that is to specifications of the form:

$$\vartheta_n^h = \sum_{k=1}^{K} \rho_{nk}\theta_{nk}^h. \tag{15}$$

Such formulations are used to allow for general specifications of the ϑ_n^h. See, for instance, Chesher & Santos Silva, (2002). If the ϑ_n^h are jointly normal distributed with an arbitrary variance-covariance matrix, (15) might refer to their Choleski decomposition. Fosgerau & Mabit (2013) show that (15) can be used to approximate any distribution of the ϑ_n^i's. This suggests that, with alternative specific constants present, the MNL cannot be distinguished from *any* other additive random utility model. That is, we cannot distinguish between an additive random utility model that has iid EV I idiosyncratic utilities and a model with *any* other random distribution of the idiosyncratic utilities when *asc*'s are present.

The significance of this result becomes clear if it is recalled that the mixed logit model considered can be any additive random utility model for a homogeneous population. With a full set of *asc*'s to account for omitted variables, no member of this class can be distinguished from MNL.

6.3 Multiple periods

It is well known that IIA does not only imply strong predictions for the consequences of introducing a new alternative in the choice set, but also for substitution behavior following a change in the attractiveness of an alternative. Indeed , from (2) it follows easily that:

$$\frac{\partial \pi_n}{\partial v_l} = I(n = l)\pi_n - \pi_n\pi_l \tag{16}$$

This is a strong prediction that appears to be testable.

If one is informed about the choice behavior of a given population in two periods, one could attempt to do this by estimating the model for both periods, assuming that MNL is valid. This gives two sets of alternative-specific constants. One could then study the changes in market shares corresponding with these changes. However, this would show a complete correspondence between prediction and reality because in both periods the predicted and actual market shares are identical. So the changes in the market shares confirm the predictions identically and no information about the validity of IIA is obtained.

If more information is available, there may be opportunities to test whether actual substitution behavior satisfies IIA. One possibility is the presence of prior information about the change in the deterministic utilities between the two periods. This may be the case in stated

choice experiments. Another is that one has panel data on individual choice behavior. The latter would allow a researcher to check if the switching from one alternative to another follows the specific pattern implied by MNL.

7 Heterogeneous actors

In this section I continue the analysis by considering heterogeneous populations. Two types of heterogeneity will be considered. I first look at random variation in the valuation of an attribute of the alternatives among the consumers. The formulation is the one commonly used in mixed, or random coefficient, logit models, although the average value of the parameter is here absorbed in the asc.

Second, I consider heterogeneity among actors linked to their observable characteristics. We are interested in what happens if there is no specification error in the deterministic parts of the utilities. With a finite number of types of actors we can estimate separate ascs for each type. This situation does not essentially change if (some aspects of) heterogeneity is related to continuous variables.

7.1 Mixed logit with random differences in the valuation of characteristics
The model refers to heterogeneity of the actors with respect to their evaluation of a characteristic, denoted as x, of the choice alternatives. The heterogeneity is unrelated to actor characteristics and therefore captured by a random variable. The deterministic part of the evaluation of x – which is common to all actors, say its mean – is absorbed in the alternative-specific constant. The deviation from that common evaluation, appears separately in the expression of the utility:

$$u_n^h = asc_n + \rho x_n \theta^h + \varepsilon_n^h \tag{17}$$

Note that the random variable θ^h takes on the same value for all choice alternatives. The multinomial logit becomes:

$$Pr_n = \int_\theta \frac{e^{asc_n + \rho x_n \theta}}{\sum_{m=1}^N e^{asc_m + \rho x_n \theta}} g(\theta)d\theta$$

$$= \int_\theta \pi_n(\theta)h(\theta)d\theta \tag{18}$$

The notation reflects that we now consider a scalar random variable. It is shown in appendix that the alternative-specific constants and the parameter ρ are not separately identified. The story is basically similar to that of the previous sections: the asc's make the loglikelihood flat in the parameter ρ. It is easy to check that the proof generalizes to situations in which the heterogeneity refers to more than one characteristic.

The situation considered here is similar to the one encountered when BLP-type models are estimated on aggregate data. Once the ascs have been determined by the contraction mapping so that observed and implied market shares are equal, the derivatives of the likelihood function with respect to all other parameters are identically equal to zero and other information is needed to estimate them.

7.2 Heterogeneity related to observed characteristics
Let the heterogeneity be indexed by $\mathbf{z}$, a vector of characteristics. Some components of $\mathbf{z}$ can be discrete, others continuous. Each household h is characterized by a value $\mathbf{z}(h)$. We don't use the notation h in this section. We refer to a particular value of $\mathbf{z}$ as a type and denote the density of a type as $f(z)$.

The situation in which there is a finite number of types is easiest to handle. All the results reached thus far hold for each of the homogeneous subpopulations consisting of a single type. As

soon as the first-order conditions for the type-specific asc's are satisfied, the likelihood function becomes flat in the other parameters of interest. Restrictions that refer to the whole population, for example that the parameters for inclusive values refer to the same sets of alternatives and have equal values for all types, therefore have no 'bite.'

If there is a continuum of types, the situation is basically similar. To analyze the issue, assume that $\mathbf{z}$ is scalar. The deterministic part of the utilities attached to the alternatives by type z actors are denoted as $v_n(z)$. These utilities are the alternative-specific constants for actors of this type. They can be estimated by local likelihood methods. We can then apply the same logic to each type. The choice probabilities for a given type z are:

$$\pi_n(z) = \frac{e^{v_n(z)}}{\sum_{m=1}^N e^{v_m(z)}}, \ n = 1 \dots N \tag{19}$$

and the expected loglikelhood $E(LL)$ is:

$$E(LL) = \int f(z)\{\sum_n ms_n(z)\ln \pi_n(z)\}dz \tag{20}$$

where ms_n denotes the expected 'market share' of alternative n for type z actors. The local likelihood of the 'asc-model' (19) estimates the market shares as choice probabilities and the two are identical if ascs are used. Therefore the same logic as before applies, once again, for any type of actor, and by implication for the continuum of types.

In practice the data will limit the researcher's ability to approximate the type-specific ascs. It is common practice when micro data are available to estimate a single alternative-specific constant for each alternative and to deal with heterogeneity related to observed characteristics of the actors through a (usually linear) parametric specification. The analysis of the present subsection makes clear that if one could estimate type-specific asc's, no room would be left for deviations from MNL like GEV or mixed logit formulations. Just like in the simple cases discussed above, it is only because of specification error in the *deterministic* part of the utilities that such deviations from iid EV I random parts of the utilities may appear to be significant. The impact of specification error for the average type has been removed through the ascs, but type-specific deviations from the average may still be misspecified if a parametric formulation is used.

8 Conclusion

This paper has shown that Hausman-McFadden and related tests for the multinomial logit model are – contrary to their intention - not informative about failure of IIA, but only about specification error in the deterministic part of the utilities attached to choice alternatives. This conclusion is reached by investigating what happens if the impact of omitted variables in the deterministic part is removed by using a full set of alternative-specific constants in a multinomial logit model. It is shown that in this context estimation on subsets of alternatives always results in the same estimates as estimation on the full set of alternatives, whereas estimation of a nested logit model is impossible because the likelihood function is flat in the parameters that embody a possible deviation from multinomial logit. Moreover, introduction of a full set of alternative-specific constants is shown to cause non-identification of any GEV-model other than the multinomial logit and of mixed logit models that have a random term that is uncorrelated with the characteristics of the alternatives. The latter result shows that the multinomial logit cannot be distinguished from any other additive random utility model. Finally, it is shown that – with a full set of alternative specific constants – coefficients for random variation in the valuation of characteristics are also unidentified, while the presence of heterogeneity related to observed characteristics of the actors does not change the results.

In summary, the results derived in this paper show that eliminating specification error from the deterministic parts of the utilities through the use of ascs makes the likelihood function

flat in parameters that embody any deviation from iid EV I random utilities. Any improvement in the likelihood that is due to a switch from MNL to a more general model that relaxes is IIA can only address specification error in the deterministic parts of the utilities that results in failure of the model to equalize predicted and observed choice frequencies for all types of actors that can be distinguished on the basis of observed characteristics.

The implication is that an extensive literature that uses maximum likelihood methods to estimate more general discrete choice models than MNL in order to avoid IIA and the associated restrictive substitution patterns is misdirected. Such models *cannot* improve on a simple MNL with a full set of alternative specific constraints for each type of actors as its only parameters and there is no empirical basis for the deviations from IIA and the more general substitution patterns suggested by such nested, cross-nested, mixed logit or any other generalization of MNL. Moreover, this is also true for probit models and any other type of additive random utility model estimated by maximum likelihood.

References
Abbe, E., Bierlaire, M., & Toledo, T. (2007). Normalization and correlation of cross-nested logit models. *Transportation Research Part B: Methodological*, *41*(7), 795–808. http://doi.org/10.1016/j.trb.2006.11.006

Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, *63*(4), 841–890.

Bresnahan, T. F., Stern, S., & Trajtenberg, M. (1997). Market segmentation and the sources of rents from innovation: Personal computers in the late 1980s. *RAND Journal of Economics*, *28*(February), S17–S44. http://doi.org/10.2307/3087454

Chesher, A., & Santos Silva, J. M. C. (2002). Taste variation in discrete choice models. *Review of Economic Studies*, *69*(1), 147–168. http://doi.org/10.1111/1467-937X.00201

Debreu, G. (1960). Review of R.D. Luce "Individual Choice Behavior." *American Economic Review*, *50*, 186–188.

Fosgerau, M., & Mabit, S. L. (2013). Easy and flexible mixture distributions. *Economics Letters*, *120*(2), 206–210. http://doi.org/10.1016/j.econlet.2013.03.050

Goldberg, P. K. (1995). Product differentiation and oligopoly in international markets: The case of the U.S. automobile industry. *Econometrica*, *63*(4), 891–951. http://doi.org/10.2307/2171803

Hahn, J., Hausman, J. A., & Lustig, J. (2017) *Specification test on mixed logit model* (cemmap working paper No. 58/17).

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica*, *46*(6), 1251–1271.

Hausman, J. A., & McFadden, D. (1984). Specification tests for the multinomial logit model. *Econometrica*. http://doi.org/10.2307/1910997

Luce, R. D. (1959). *Individual choice behavior: A theoretical analysis*. New York: Wiley.

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics* (pp. 105–142).

McFadden, D. (1978). Modelling the choice of residential location. In A. Karqvist, L. Lundqvist, & F. Snickars (Eds.), *Spatial Interaction Theory and Planning Models* (pp. 75–96). Amsterdam: Elsevier. Retrieved from http://cowles.econ.yale.edu/P/cd/d04b/d0477.pdf

McFadden, D., & Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, *15*(November 1998), 447–470. http://doi.org/10.1002/1099-1255(200009/10)15:5<447::AID-JAE570>3.0.CO;2-1

McFadden, D., Train, K. E., & Tye, W. B. (1981). An application of diagnostic tests for the

independence from irrelevant alternatives property of the multinomial logit model. *Transportation Research Record.*

Small, K. A. (1987). A discrete choice model for ordered alternatives. *Econometrica*, *55*(2), 409–424. http://doi.org/10.2307/1913243

Small, K. A., & Hsiao, C. (1985). Multinomial logit specification tests. *International Economic Reviewc Review*, *26*(3), 619–627. http://doi.org/10.2307/2526707

Train, K. E. (2003). *Discrete Choice Methods with Simulation. Cambridge University Press.* Cambridge: CUP. Retrieved from http://ebooks.cambridge.org/ref/id/CBO9780511753930

Appendix
A Estimation of the nested logit model with alternative-specific constants

A1 The first order condition for the asc's and some implications

The model is (10) and we want to recover the utilities as alternative specific constants. So we In substitute $asc_n = v_n$ and write the loglikelihood ll_n of a choice for alternative n:

$$ll_n = \frac{asc_n}{\lambda_{k(n)}} + \left(\lambda_{k(n)} - 1\right) \ln\left(\sum_{m \in B_{k(n)}} e^{asc_m/\lambda_{k(n)}}\right) - \ln\left(\sum_{k=1}^{K}\left(\sum_{m \in B_k} e^{asc_m/\lambda_k}\right)\right) \tag{A1}$$

Differentiation with respect to the asc's gives:

$$\frac{\partial ll_n}{\partial asc_l} = I(l = n)\frac{1}{\lambda_{k(l)}} + I\left(l \in B_{k(n)}\right)\left(\lambda_{k(l)} - 1\right)\pi_{l|k(l)}\frac{1}{\lambda_{k(l)}} + \pi_l \tag{A2}$$

where $I(.)$ is an indicator function that takes on the value 1 if the expression that appears on the . is true and 0 otherwise, and $\pi_{l|k(l)}$ is the probability that alternative l will be chosen conditional on the choice for the nest to which it belongs:

$$\pi_{l|k(l)} = \frac{e^{asc_l/\lambda_{k(l)}}}{\sum_{m \in B_{k(l)}} e^{asc_m/\lambda_{k(l)}}} \tag{A3}$$

The total loglikelihood LL is the sum of the individual loglikelihoods over all observations:

$$LL = \sum_{n=1}^{N} f_n ll_n \tag{A4}$$

where f_n denotes the number of observed choices for alternative n. It follows then that:

$$\frac{\partial LL}{\partial asc_l} = f_l \frac{1}{\lambda_{k(l)}} + \frac{\lambda_{k(l)} - 1}{\lambda_{k(l)}} F_{k(l)} \pi_{l|k(l)} + F \pi_l \tag{A5}$$

where F_k denotes the number of observed choices for alternatives belonging to nest k (i.e. $F_k = \sum_{l \in B_k} f_l$) and F is the total number of observations.

Setting $\frac{\partial LL}{\partial asc_l}$ equal to zero gives:

$$\frac{1}{\lambda_{k(l)}}\left(f_l - F_k \pi_{l|k(l)}\right) + \left(F \pi_l - F_k \pi_{l|k(l)}\right) = 0 \tag{A6}$$

If we sum this expression over all alternatives in nest k, the first term becomes:

$$\frac{1}{\lambda_{k(l)}}\left(F_k - F_k \sum_{l \in B_{l(l)}} \pi_{l|k(l)}\right) \tag{A7}$$

and since the sum of the conditional choice probabilities equals 1, this is equal to zero. We conclude therefore that the sum of the second term over all alternatives in nest k must also be equal to zero:

$$F \pi_{B_k} - F_k = 0 \tag{A8}$$

where π_{B_k} is the probability that an alternative in nest k will be chosen (i.e. $\pi_{B_k} = \sum_{l \in B_k} \pi_l$. Multiplying both sides of (A8) with $\pi_{l|k(l)}$ we get:

$$F \pi_l - F_k \pi_{l|k(l)} = 0 \tag{A9}$$

and substitution of this result in (A6) gives us finally:

$$f_l - F_k \pi_{l|k(l)} = 0 \tag{A10}$$

A2 The first-order condition of the coefficients for the inclusive values

Differentiation of the likelihood in (A1) with respect to a coefficient for an inclusive value gives:

$$\frac{\partial ll_n}{\partial \lambda_i} =$$

$$I(k(n) = i)\left(-\frac{asc_n}{\lambda_i^2}\right) +$$

$$I(k(n) = i) \left\{ \ln\left(\sum_{m \in B_i} e^{asc_m/\lambda_i}\right) + (\lambda_i - 1)\left(\sum_{m \in B_i} \pi_{m|i}\left(-\frac{asc_n}{\lambda_i^2}\right)\right) \right\} -$$

$$\pi_{B_i} \left\{ \ln\left(\sum_{m \in B_i} e^{asc_m/\lambda_i}\right) + \lambda_i \left(\sum_{m \in B_i} \pi_{m|i}\left(-\frac{asc_n}{\lambda_i^2}\right)\right) \right\} \tag{A11}$$

Summation over all observations gives:

$$\frac{\partial LL}{\partial \lambda_i} = \sum_{m \in B_i} f_m\left(-\frac{asc_m}{\lambda_i^2}\right) + F_i \left\{ \ln\left(\sum_{m \in B_i} e^{asc_m/\lambda_i}\right) + (\lambda_i - 1)\left(\sum_{m \in B_i} \pi_{m|i}\left(-\frac{asc_m}{\lambda_i^2}\right)\right) \right\} -$$

$$F \, \pi_{B_i} \left\{ \ln\left(\sum_{m \in B_i} e^{asc_m/\lambda_i}\right) + \lambda_i \left(\sum_{m \in B_i} \pi_{m|i}\left(-\frac{asc_n}{\lambda_i^2}\right)\right) \right\} \tag{A12}$$

Rearranging terms gives:

$$\frac{\partial LL}{\partial \lambda_i} =$$

$$\sum_{m \in B_i} [f_m - F_i \pi_{m|i}] \left(-\frac{asc_m}{\lambda_i^2}\right) +$$

$$[F_i - F \, \pi_{B_i}] \left\{ \ln\left(\sum_{m \in B_i} e^{asc_m/\lambda_i}\right) + \lambda_i \left(\sum_{m \in B_i} \pi_{m|i}\left(-\frac{asc_n}{\lambda_i^2}\right)\right) \right\} \tag{A13}$$

The two expressions in square brackets are equal to 0. For the first one this follows from (A10). For the second one from (A8). Hence $\frac{\partial LL}{\partial \lambda_i}$ is identically zero once the first order conditions for the asc's are satisfied.

B Estimation of GEV models with alternative-specific constants

We consider a GEV model where the utilities are specified as alternative-specific constants. GEV models are characterized by a generator function $G(y; \theta)$ which is homogeneous of degree 1 in y, an N-dimensional vector of arguments and satisfies a number of other properties listed in McFadden (1978). The symbol θ denotes the parameters of this function. Although we will proceed below as if it is a scalar to keep the notation simple, it can be a vector of arbitrary dimension. For instance, with a nested logit model the parameters are those of the inclusive values. The homogeneity of G implies $G = \sum_n y_n G_n$, where G_n denotes the first partial derivative of G with respect to its n-th element. The choice probabilities can be written as:

$$\pi_n = \frac{e^{asc_n} G_n(e^{asc}, \theta)}{G(e^{asc}, \theta)}, \quad n = 1 \dots N. \tag{B1}$$

In this equation e^{asc} is the vector with n-th element e^{asc_n}. The log likelihood ll_n of a choice for the n-th alternative is:

$$ll_n = asc_n + \ln G_n - \ln \sum_{m=1}^{N} e^{asc_m} G_m \tag{B2}$$

Taking the first derivative with respect to θ gives:

$$\frac{\partial ll_n}{\partial \theta} = \frac{G_{n\theta}}{G_n} - \sum_{m=1}^{N} \pi_m \frac{G_{m\theta}}{G_m} \tag{B3}$$

In this equation $G_{n\theta}$ denotes the partial derivative of G_n with respect to θ. Defining the total loglikelihood similar as in the previous part of the Appendix, we find:

$$\frac{\partial LL}{\partial \theta} = \sum_{m=1}^{N} (f_m - F\pi_m) \frac{G_{m\theta}}{G_m} \tag{B4}$$

where f_m denotes the number of observed choices for alternative m and F the total number of observed choices.

Now consider the first derivative of ll_n with respect to asc_l:

$$\frac{\partial ll_n}{\partial asc_l} = I(l = n) + \frac{e^{asc_l}G_{nl}}{G_n} - \pi_l - \sum_{m=1}^N \pi_m \frac{e^{asc_l}G_{ml}}{G_m} \tag{B5}$$

where G_{nl} denotes the partial derivative of G_n with respect to e^{asc_l}. It follows then that:

$$\frac{\partial LL}{\partial asc_l} = f_l + \sum_{m=1}^N f_m \frac{e^{asc_l}G_{ml}}{G_m} - F\pi_l - F\sum_{m=1}^N \pi_m \frac{e^{asc_l}G_{ml}}{G_m} \tag{B6}$$

Now substitute (B1) into the last term and rewrite it as:

$$F\sum_{m=1}^N \pi_m \frac{e^{asc_l}G_{ml}}{G_m} = \frac{Fe^{asc_l}}{G}\sum_{m=1}^N e^{asc_m}G_{ml} \tag{B7}$$

Since G is homogeneous of degree 1, its partial derivatives will be homogeneous of degree 0. This implies that the sum on the right-hand side of (B7) will be equal to zero. Hence we can simplify (B6) to:

$$\frac{\partial LL}{\partial asc_l} = (f_l - F\pi_l) + \sum_{m=1}^N f_m \frac{e^{asc_l}G_{ml}}{G_m} \tag{B8}$$

This must be equal to zero. This is the case if we set the ascs so that $f_l - F\pi_l = 0$. To see that this suffices, start by observing that the first term on the right-hand side of (B8) then equals 0. This is also true for the second term: if we substitute $f_m = F\pi_m$ into it, it can be rewritten as:

$$\sum_{m=1}^N f_m \frac{e^{asc_l}G_{ml}}{G_m} = \frac{Fe^{asc_l}}{G}\sum_{m=1}^N e^{asc_m}G_{ml} \tag{B9}$$

The sum on the right-hand side is equal to 0 because the partial derivatives of G are homogeneous of degree 0.

If $f_m = F\pi_l, l = 1 \ldots N$, (B4) implies that $\frac{\partial LL}{\partial \theta}$ will be equal to 0. We thus find the same result as for the specific case of the nested logit: if the first-order conditions with respect to the choice of the ascs are satisfied, the first-order condition with respect to the choice of *any* parameter of the generator function of the generalized extreme value distribution is also automatically satisfied.

C Estimation of mixed logit models with alternative-specific random terms

The model is specified in subsection 4.2. We begin with the computation of the partial derivatives of $ll_n = \ln Pr_n$ w.r.t. asc_l and ρ_l:

$$\frac{\partial ll_n}{\partial asc_l} = \frac{1}{Pr_n}\int_\theta \{\pi_n(\boldsymbol{\theta})I(n = l) - \pi_n(\boldsymbol{\theta})\pi_l(\boldsymbol{\theta})\}g(\boldsymbol{\theta})d\boldsymbol{\theta}$$

$$= I(n = l) - \frac{1}{Pr_n}\int_\theta \pi_n(\boldsymbol{\theta})\pi_l(\boldsymbol{\theta})g(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{C1}$$

$$\frac{\partial ll_n}{\partial \rho_l} = \frac{1}{Pr_n}\int_\theta \{\pi_n(\boldsymbol{\theta})I(n = l) - \pi_n(\boldsymbol{\theta})\pi_l(\boldsymbol{\theta})\}\theta_l g(\boldsymbol{\theta})d\boldsymbol{\theta}$$

$$= I(n = l)\frac{1}{Pr_n}\int_\theta \pi_n(\boldsymbol{\theta})\theta_l g(\boldsymbol{\theta})d\boldsymbol{\theta} - \frac{1}{Pr_n}\int_\theta \pi_n(\boldsymbol{\theta})\pi_l(\boldsymbol{\theta})\theta_l g(\boldsymbol{\theta})d\boldsymbol{\theta} \tag{C2}$$

For the likelihood of the whole sample we must have:

$$\left(\frac{\partial LL}{\partial asc_l} = 0 \Rightarrow\right) f_l - \sum_n \frac{f_n}{Pr_n}\int_\theta \pi_n(\boldsymbol{\theta})\pi_l(\boldsymbol{\theta})g(\boldsymbol{\theta})d\boldsymbol{\theta} = 0 \tag{C3}$$

$$\left(\frac{\partial LL}{\partial \rho_l} = 0 \Rightarrow\right) \frac{f_l}{Pr_l}\int_\theta \pi_l(\boldsymbol{\theta})\theta_l g(\boldsymbol{\theta})d\boldsymbol{\theta} - \sum_n \frac{f_n}{Pr_n}\int_\theta \pi_n(\boldsymbol{\theta})\pi_l(\boldsymbol{\theta})g(\boldsymbol{\theta})d\boldsymbol{\theta} = 0 \tag{C4}$$

Now choose the alternative-specific constants so that $Pr_n = f_n/F$, n=1…N. Substitution in (C3) then gives:

$$f_l - F\int_\theta \left(\sum_n \pi_n(\boldsymbol{\theta})\right)\pi_l(\boldsymbol{\theta})g(\boldsymbol{\theta})d\boldsymbol{\theta} = 0 \tag{C5}$$

which is identically true since $\sum_n \pi_n(\boldsymbol{\theta}) = 1$. Hence this solves (C3).
Substitution in (C4) gives:

$$\int_\theta \pi_l(\boldsymbol{\theta})\theta_l g(\boldsymbol{\theta})d\boldsymbol{\theta} - \int_\theta \left(\sum_n \pi_n(\boldsymbol{\theta})\right)\pi_l(\boldsymbol{\theta})g(\boldsymbol{\theta})d\boldsymbol{\theta} = 0 \tag{C6}$$

which is also identically true.

Hence setting the alternative-specific constants so that the choice probabilities equal the relative frequencies solves the first order conditions for the alternative-specific constants asc_n, $n = 1 \dots N$, as well as those for the heterogeneity parameters ρ_n, $n = 1 \dots N$. The latter can thus be given any desired value.

D Estimation of mixed logit models with random variation in attribute valuation

The model is given by (17). We proceed similarly as in the previous appendix and begin with the computation of the partial derivatives of $ll_n = \ln Pr_n$ w.r.t. asc_l and ρ_l:

$$\frac{\partial ll_n}{\partial asc_l} = \frac{1}{Pr_n} \int_\theta \{\pi_n(\theta)I(n = l) - \pi_n(\theta)\pi_l(\theta)\}g(\theta)d\theta$$

$$= I(n = l) - \frac{1}{Pr_n} \int_\theta \pi_n(\theta)\pi_l(\theta)g(\theta)d\theta \tag{D1}$$

$$\frac{\partial ll_n}{\partial \rho} = \frac{1}{Pr_n} \int_\theta \{\pi_n(\theta)x_n\theta I(n = l) - \pi_n(\theta)\sum_l \pi_l(\theta)x_l\}\theta g(\theta)d\theta$$

$$= I(n = l)\frac{1}{Pr_n} \int_\theta \pi_n(\theta)x_n\theta g(\theta)d\theta - \frac{1}{Pr_n} \int_\theta \pi_n(\theta)(\sum_l \pi_l(\theta)x_l)\theta g(\theta)d\theta \tag{D2}$$

For the likelihood of the whole sample we must have:

$$\left(\frac{\partial LL}{\partial asc_l} = 0 \Rightarrow\right) f_l - \sum_n \frac{f_n}{Pr_n} \int_\theta \pi_n(\theta)\pi_l(\theta)g(\theta)d\theta = 0 \tag{D3}$$

$$\left(\frac{\partial LL}{\partial \rho} = 0 \Rightarrow\right) \sum_l \left\{\frac{f_l}{Pr_l} \int_\theta \pi_l(\theta)x_l\theta g(\theta)d\theta - \sum_n \frac{f_n}{Pr_n} \int_\theta \pi_l(\theta)\pi_n(\theta)x_n\theta g(\theta)d\theta\right\} = 0 \tag{D4}$$

Now choose the alternative-specific constants so that $Pr_n = f_n/F$, n=1…N. Substitution in (D3) then gives:

$$f_l - F \int_\theta (\sum_n \pi_n(\theta))\pi_l(\theta)g(\theta)d\theta = 0 \tag{D5}$$

which is identically true since $\sum_n \pi_n(\theta) = 1$. Hence this solves (D3).
Substitution in (D4) gives:

$$\sum_l \left\{\int_\theta \pi_l(\theta)x_l\theta g(\theta)d\theta - \sum_n \int_\theta \pi_l(\theta)\pi_n(\theta)x_n\theta g(\theta)d\theta\right\} = 0 \tag{D6}$$

The left-hand side can be rewritten as:

$$\sum_l \left\{\int_\theta \pi_l(\theta)x_l\theta g(\theta)d\theta\right\} - \sum_n \int_\theta (\sum_l \pi_l(\theta))\pi_n(\theta)x_n\theta g(\theta)d\theta \tag{D7}$$

Which is identically zero since $\sum_l \pi_l(\theta) = 1$.