Specification Test on Mixed Logit Models

Jinyong Hahn Jerry Hausman Josh Lustig

UCLA MIT CRA

December 10, 2017

Abstract

This paper proposes a specification test of the mixed logit models, by generalizing Hausman and McFadden's (1984) test. We generalize the test even further by considering a model developed by Berry, Levinsohn and Pakes (1995).

1 Introduction

Multinomial choice models have become an important model in demand estimation. The model can parsimoniously characterize the demand system by allowing the number of parameters to be substantially smaller than the number of products. In this literature, it is a common practice to adopt a mixed logit specification, probably for the purpose of relaxing the Independence of Irrelevant Alternatives (IIA) properties. However, the IIA property still holds at the individual level. A model specification which relaxes the individual IIA assumption is formulated and estimated by Burda, Harding and Hausman (2008).

The logit specification, despite such limitations, provides much computational convenience, which naturally prompts for a specification test. A specification test for the multinomial logit model was addressed by Hausman and McFadden (1984), who proposed a variation of the Hausman (1978) test. We note that the specification test does not exist for mixed logit models.

It has been recognized for many years that an important problem with the multinomial logit model is the Independence of Irrelevant Alternatives (IIA) property. The IIA property implies that the ratio of the probabilities of choosing any two alternatives is independent of the attributes of any other alternative in the choice set. Debreu (1960) gave an early discussion about the implausibility of the IIA assumption. Models that have the IIA property do not allow for different degrees of substitution or complementarity among the choices.[1] Indeed, Hausman (1975) demonstrate that IIA requires all cross-price elasticities for a given product are identical, a seemingly implausible assumption for differentiated product demand models.

An early justification for use of the multinomial logit model with the IIA property was that when estimated on individual data, aggregate predictions did not have the IIA property. The Hausman–McFadden (1984) test allowed for a test of the underlying foundational IIA assumption. No known property was demonstrated in the literature that if the IIA property did not hold at the individual level, it "cancelled out" at the aggregate level in terms of estimating the correct price elasticities. More recently, models which allowed heterogenous preferences have become widely used. Again, as we demonstrate in this paper and has been recognized, many heterogenous preference models impose IIA at the individual level. Again, claims have been made that when used at the aggregate level that the IIA assumption has only limited relevance. We explore those claims in this paper and provide a specification test that allow for a determination whether use of the IIA property at the individual level leads to inconsistent estimates at the aggregate level, when the IIA property does not hold true.

Thus, the purpose of this paper is to fill the gap in the literature by developing a generalization of the Hausman and McFadden's (1984) specification test. We consider two variants of the test. In the first case, we consider the usual mixed logit model where the logit model parameters are assumed to be random coefficients independent of the explanatory variables. As in Hausman and McFadden (1984), we consider estimating the model coefficients after removing an alternative from the choice set, and comparing the new parameter estimates with the original estimates. Since the mixed logit model assumes IIA at the individual level, the specification test should have good power properties. In the second case, we consider the variant of the mixed logit model considered

[1] In the literature, this property is often called the "red bus-blue bus" problem. The busses should have different substitution properties between themselves than substitution with say Lyft or the Metro. Hausman often thought this example was too extreme, but as with black swans discovered in Western Australia, he has discovered that two taxpayer funded bus operations in Santa Monica, CA, with significant overlap in some of their routes, use blue and red buses.

by Berry, Levinsohn and Pakes (1995, BLP hereafter), in which a coefficient and its associated variable may exhibit endogeneity.

2 Mixed Logit Model

In this section, we consider a typical mixed logit choice model, and develop a specification test in the spirit of Hausman and McFadden (1984). We compare two parameter estimates. The first one is the maximum likelihood estimator (MLE) for the original model. The second one is the MLE for the implied model where we remove an alternative from the choice set. The removal of an alternative produces a sample selection problem, which we control by using Bayes theorem. The resultant likelihood for the restricted model turns out to be very intuitive. Because the original MLE is the efficient estimator, the comparison of the two estimates validates the straightforward formula as derived in Hausman (1978).

We start with a standard model where the individual utility takes the form

$$U_{i,j} = \gamma_i' x_j + \epsilon_{i,j}, \quad j = 1, \ldots, J; \quad i = 1, \ldots, N. \tag{1}$$

We assume that $\epsilon_{i,j}$ are independent and identically distributed extreme value random variables. The γ_i is the random coefficient which allows for deviation from the textbook logit model. We assume that the γ_i are i.i.d. with distribution $f(\gamma|\,\theta)$ parameterized by some θ. Typically, we assume that it is drawn from a multivariate normal distribution, although we leave $f(\gamma|\,\theta)$ as an arbitrary distribution throughout the theory part of the paper. Some of the components of γ_i may be allowed to be nonrandom. Indeed, by allowing $f(\gamma|\,\theta)$ to be a mixture distribution, a very flexible distribution can be used as discussed in e.g. Burda, Harding and Hausman (2008).

We can easily see that the model above nests the typical model used in demand estimation

$$U_{i,j} = \beta_i' \mathbf{x}_j - \alpha P_j + \epsilon_{i,j}, \tag{2}$$

where P_j is price and $\mathbf{x}_j$ are non-price product characteristics. For simplicity, we assume that the x_j in (1) is not individual specific, i.e., it does not have the i subscript. This assumption has the benefit of having a similar notation as in the analysis of the BLP afterwards, but in the context of a specification test of the mixed logit model with individual data, it is not necessary.

Before we describe our own test, it may be helpful to review the intuition underlying Hausman and McFadden's (1984) test. Assume that γ_i is fixed at γ in model (1). For simplicity, assume that $J = 3$. For the full model, we see that

$$\Pr\left(d_{ij} = 1 \mid C\right) = \frac{\exp\left(\gamma' x_j\right)}{\exp\left(\gamma' x_1\right) + \exp\left(\gamma' x_2\right) + \exp\left(\gamma' x_3\right)}, \tag{3}$$

where d_{ij} denotes the binary indicator that takes the value 1 if the ith individual chooses the jth alternative. Here, the notation C denotes the original choice set. Now, if we remove the third alternative, the IIA implies that

$$\Pr\left(d_{ij} = 1 \mid C^R\right) = \frac{\exp\left(\gamma' x_j\right)}{\exp\left(\gamma' x_1\right) + \exp\left(\gamma' x_2\right)}, \tag{4}$$

where C^R denotes the restricted choice set that removes the third alternative. The Hausman and McFadden (1984) test compares the MLE for the full sample using the specification (3), and the MLE for the subsample with $y_{i3} = 0$ using the specification (4). The IIA assumption of the logit model follows from equations (3) and (4), where the ratios of the probabilities for $d_{i1} = 1$ and $d_{i2} = 1$ are the same in both equations.

We now consider a generalization of this idea to the mixed logit model. In this context, developing the likelihood for the subsample requires controlling for selection. For this purpose, consider removing the "outside good" in the example (2). Individuals who choose the outside good have preferences that are different from individuals who don't choose the outside good. In the current setting, individuals who have weak preferences for the two non-price characteristics (i.e., individuals with low or negative β) are most likely to choose the outside good. Therefore, if we remove the outside good from the choice set and estimate the model using only individuals who didn't chose the outside good, the parameter estimates will overstate the individuals' willingness to pay for non-price characteristics. The selection problem arises from the presence of random coefficients $f\left(\gamma \mid \theta\right)$.

We now consider removing the last alternative in the mixed logit specification (1). For simplicity, we will assume that $J = 3$. For individuals with γ_i, the probabilities that the three options are chosen are given by

$$p_j\left(\gamma_i\right) \equiv \Pr\left(d_{ij} = 1 \mid \gamma_i\right) = \frac{\exp\left(\gamma_i' x_j\right)}{\exp\left(\gamma_i' x_1\right) + \exp\left(\gamma_i' x_2\right) + \exp\left(\gamma_i' x_3\right)}, \tag{5}$$

and the (unconditional) likelihood is the integrated version with respect to $f(\gamma|\theta)$, i.e.,

$$P(j;\theta) \equiv \Pr(d_{ij}=1|\theta) = \int \frac{\exp(\gamma' x_j)}{\exp(\gamma' x_1)+\exp(\gamma' x_2)+\exp(\gamma' x_3)} f(\gamma|\theta)\, d\gamma. \tag{6}$$

Note that the IIA holds at the individual level in equation (5). We consider removing the last alternative, with the restricted choice set C^R consisting of $j=1,2$. Note that

$$\Pr\left(C^R\middle|\gamma_i\right) = \frac{\exp(\gamma_i' x_1)+\exp(\gamma_i' x_2)}{\exp(\gamma_i' x_1)+\exp(\gamma_i' x_2)+\exp(\gamma_i' x_3)}, \tag{7}$$

where $\Pr\left(C^R\middle|\gamma_i\right)$ denotes the probability that the restrictive choice set C^R is chosen. The IIA implies that

$$\Pr\left(d_{i1}=1\middle|\gamma_i, C^R\right) = \frac{\exp(\gamma_i' x_1)}{\exp(\gamma_i' x_1)+\exp(\gamma_i' x_2)}.$$

It follows that

$$\begin{aligned}\Pr\left(d_{i1}=1\middle|C^R\right) &= \int \Pr\left(y_{i1}=1\middle|\gamma, C^R\right) f\left(\gamma\middle|C^R,\theta\right) d\gamma \\ &= \int \frac{\exp(\gamma_i' x_1)}{\exp(\gamma_i' x_1)+\exp(\gamma_i' x_2)} f\left(\gamma\middle|C^R,\theta\right) d\gamma, \end{aligned} \tag{8}$$

where $f\left(\gamma\middle|C^R,\theta\right)$ denotes the conditional density of γ for the subsample of individuals that chose the alternatives in C^R. By Bayes rule, we have

$$\begin{aligned} f\left(\gamma\middle|C^R\right) &= \frac{\Pr\left(C^R\middle|\gamma\right) f(\gamma|\theta)}{\int \Pr\left(C^R|\gamma\right) f(\gamma|\theta)\, d\gamma} \\ &= \frac{1}{\Gamma\left(C^R|\theta\right)} \frac{\exp(\gamma' x_1)+\exp(\gamma' x_2)}{\exp(\gamma' x_1)+\exp(\gamma' x_2)+\exp(\gamma' x_3)} f(\gamma|\theta), \end{aligned} \tag{9}$$

where we write

$$\begin{aligned} \Gamma\left(C^R\middle|\theta\right) &\equiv \int \Pr\left(C^R\middle|\gamma\right) f(\gamma|\theta)\, d\gamma \\ &= \int \frac{\exp(\gamma' x_1)+\exp(\gamma' x_2)}{\exp(\gamma' x_1)+\exp(\gamma' x_2)+\exp(\gamma' x_3)} f(\gamma|\theta)\, d\gamma. \end{aligned}$$

Combining (6), (8) and (9), we obtain

$$\begin{aligned} \Pr\left(d_{i1}=1\middle|C^R\right) &= \frac{1}{\Gamma\left(C^R|\theta\right)} \int \frac{\exp(\gamma' x_1)}{\exp(\gamma' x_1)+\exp(\gamma' x_2)+\exp(\gamma' x_3)} f(\gamma|\theta)\, d\gamma \\ &= \frac{P(1;\theta)}{\Gamma\left(C^R|\theta\right)}. \end{aligned}$$

Likewise, we obtain

$$\Pr\left(d_{i2}=1\middle|\, C^R\right)=\frac{P\left(2;\theta\right)}{\Gamma\left(C^R\middle|\,\theta\right)}.$$

It is straightforward to show that the result generalizes in a straightforward manner to the case with arbitrary J and C^R. In order to characterize the likelihoods, it is convenient to define a random variable $y_i=1$ if $d_{ij}=1$. Also, let $z_i=1$ if the agent i chooses an option in C^R, and 0 otherwise. Then the MLE $\widehat{\theta}_1$ based on the full sample solves

$$\max_{\theta}\sum_{i=1}^{N}\log P\left(y_i;\theta\right), \tag{10}$$

where $P\left(j;\theta\right)$ is defined in (6). The MLE $\widehat{\theta}_2$ based on the subsample after certain choices are removed solves

$$\max_{\theta}\sum_{i=1}^{N}z_i\log\frac{P\left(y_i;\theta\right)}{\Gamma\left(C^R\middle|\,\theta\right)}. \tag{11}$$

Under correct specification and standard regularity conditions, we can see that both estimator are consistent and asymptotically normal, and their asymptotic variances can be consistently estimated by

$$\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{\partial\log P\left(y_i;\widehat{\theta}_1\right)}{\partial\theta}\right)\left(\frac{\partial\log P\left(y_i;\widehat{\theta}_1\right)}{\partial\theta}\right)'\right]^{-1} \tag{12}$$

and

$$\left[\frac{1}{N}\sum_{i=1}^{N}z_i\left(\frac{\partial\log\frac{P\left(y_i;\widehat{\theta}_2\right)}{\Gamma\left(C^R\middle|\,\widehat{\theta}_2\right)}}{\partial\theta}\right)\left(\frac{\partial\log\frac{P\left(y_i;\widehat{\theta}_2\right)}{\Gamma\left(C^R\middle|\,\widehat{\theta}_2\right)}}{\partial\theta}\right)'\right]^{-1}. \tag{13}$$

Because $\widehat{\theta}_1$ is efficient relative to $\widehat{\theta}_2$, the Hausman test statistic takes the usual form.

Comparison of (10) and (11) does indeed make a natural generalization of Hausman and McFadden (1984), which can be understood by considering a simple case without any random coefficient, i.e., the case where $\gamma_i=\theta$. If so, we obtain

$$\begin{aligned}
P\left(j\middle|\,\theta\right)&=\frac{\exp\left(\theta'x_j\right)}{\exp\left(\theta'x_1\right)+\exp\left(\theta'x_2\right)+\exp\left(\theta'x_3\right)}\\
\Gamma\left(C^R\middle|\,\theta\right)&=\frac{\exp\left(\theta'x_1\right)+\exp\left(\theta'x_2\right)}{\exp\left(\theta'x_1\right)+\exp\left(\theta'x_2\right)+\exp\left(\theta'x_3\right)}\\
\frac{P\left(j\middle|\,\theta\right)}{\Gamma\left(C^R\middle|\,\theta\right)}&=\frac{\exp\left(\theta'x_j\right)}{\exp\left(\theta'x_1\right)+\exp\left(\theta'x_2\right)}
\end{aligned}$$

Therefore, the counterpart of (11) indeed reflects the IIA (4).

3 Extension to BLP

In this section, we generalize the idea developed in Section 2 to deal with the complications in BLP. We develop counterparts of $\widehat{\theta}_1$ and $\widehat{\theta}_2$, and discuss how they can be compared. We will call $\widehat{\theta}_1$ and $\widehat{\theta}_2$ the first and second step estimators.

3.1 Characterization of $\widehat{\theta}_1$

Characterization of the first step estimator is relatively straightforward, because it only requires description of the BLP model. We do need to be a little bit careful in describing the asymptotic framework. The BLP typically starts with the utility We start with the utility

$$U_{i,j} = \mathbf{x}_j \beta_i - p_j \alpha + \xi_j + \epsilon_{i,j} = \gamma_i' w_j + \xi_j + \epsilon_{i,j}, \tag{14}$$

where $\epsilon_{i,j}$ is i.i.d. extreme value distribution $j = 1, \ldots, J$. The market share s_j is then

$$s_j = \int p\left(y_j \middle| w, \gamma, \xi\right) f\left(\gamma \middle| \theta\right) d\gamma, \tag{15}$$

where w denotes the collection of $\mathbf{x}$'s and p's, and the $f\left(\gamma \middle| \theta\right)$ denotes the density of $\gamma = (\beta, \alpha)$ indexed by θ. It is assumed that there is an instrument such that[2]

$$E\left[z\xi_j\right] = 0, \quad j = 1, \ldots, J. \tag{16}$$

Using the contraction mapping discussed in BLP, we can write

$$\xi_j = g_j\left(s^0, w, f\left(\cdot \middle| \theta\right)\right), \tag{17}$$

where s^0 denotes the vector of shares in the population. Letting F_θ denote the distribution of γ, we may write the moment restriction

$$E\left[z g_j\left(s^0, w, F_\theta\right)\right] = 0, \quad j = 1, \ldots, J \tag{18}$$

based on which we can estimate θ.[3]

[2] In BLP, the instrument is in fact a function $H_j(\mathbf{z})$ of the conditioning variable $\mathbf{z}$, where $\mathbf{z}$ is from the conditional moment restriction $E\left[\xi_j \middle| \mathbf{z}\right] = 0$. We avoid notational complication by working with the instrument z_j itself.

[3] Here, the s^0 denotes the true vector of shares, but in practice we use the estimated vector of shares s^n instead. The difference does not result in different asymptotic distribution as long as $n, T \to \infty$ at an appropriate rate, which can be shown by a textbook-level analysis. See Appendix B.

In order to understand the moment (18) in a convenient asymptotic framework, we use inter-market variation and work with

$$E\left[z_t g_j\left(s_t^0, w_t, F_\theta\right)\right] = 0 \quad , \quad j = 1, \ldots, J; t = 1, \ldots, T \tag{19}$$

for $\xi_{j,t} = g_j\left(s_t^0, w_t, F_\theta\right)$. In terms of asymptotics, we assume that J is fixed while $T \to \infty$.[4] This approach leads to the characterization of the first step estimator to be the solution to the sample counterpart[5] of the (18) in the following form:

$$0 = \frac{1}{T} \sum_{t=1}^{T} \sum_{j} z_{j,t} g_j\left(s_t^0, w_t, F_{\widehat{\theta}_1}\right). \tag{20}$$

Here, the $z_{j,t}$ denotes an arbitrary transformation of z_t.

3.2 Characterization of $\widehat{\theta}_2$

We now consider implementation of the second step, and consider estimation of θ after removing an alternative. Roughly speaking, the implementation of the second step consists of the following: First, we define the restricted choice set C^R as before. We note that by Bayes rule, this approach is equivalent to using

$$\frac{\Pr\left(C^R \middle| \gamma\right) f_\theta(\gamma)}{\int \Pr\left(C^R \middle| \gamma\right) f_\theta(\gamma)\, d\gamma} \tag{21}$$

as the density of γ, instead of $f_\theta(\gamma)$. Let F_θ^R denote such a distribution. Note that the F_θ^R in fact depends on w_t, so it should in principle indexed by t as well, although we suppress it here for notational simplicity.

With the restricted choice set, we need to redefine the vector of market shares s_t^R. For simplicity, we assume that the first J_1 alternatives constitute the restricted choice set, and the last $J - J_1$ alternatives are removed. The s_t^R is then the J_1-dimensional vector which is obtained by choosing the first J_1 elements of s_t^0 and dividing each of them by the sum of the J_1 elements. For example, suppose that there are four choices in the original choice set, i.e., s_t^0 is a four-dimensional vector. Suppose that the last choice had a market share equal to 20%. Suppose that C^R consists of the first three choices. Then s_t^R is a three-dimensional vector obtained by dividing the first three

[4] Our asymptotics reflects Berry and Haile's (2014) result. See Appendix A.

[5] Here $z_{j,t}$ corresponds to $H_j(z) T(z_j)$ in BLP (p. 857).

components of s_t^0 by $100\%-20\% = 80\%$. We then use $g_j\left(s_t^R, w_t, F_\theta^R\right)$ in (20).[6] The rest is identical to the first step.

We now provide the details of implementation. We recognize two features that distinguishes the current models of the BLP framework. First, the discussion in the previous section makes it clear that the second step requires the counterpart of g_j in (18) be based on the conditional the conditional density of γ for the subsample of individuals that chose the alternatives in C^R. Second, we have additional ξ in each market, which is fixed in a given market and plays a role of a parameter in each market. Therefore, implementation of (21) requires careful re-examination of our steps in the previous section.

In order not to complicate notations unnecessarily, we proceed as before and only consider the simple case $J = 3$, where we remove the third alternative. Writing[7]

$$U_{i,j,t} = \gamma_i' w_{j,t} + \xi_{j,t} + \epsilon_{i,j,t},$$

we obtain the counterparts of (6) and (7)

$$\begin{aligned} P_t\left(j;\theta,\xi_t,w_t\right) &\equiv \Pr\left(d_{ijt}=1\middle|\,\theta,\xi_t,w_t\right) \\ &= \int \frac{\exp\left(\gamma' w_{j,t}+\xi_{j,t}\right)}{\exp\left(\gamma' w_{1,t}+\xi_{1,t}\right)+\exp\left(\gamma' w_{2,t}+\xi_{2,t}\right)+\exp\left(\gamma' w_{3,t}+\xi_{3,t}\right)} f\left(\gamma\middle|\,\theta\right)d\gamma, \end{aligned} \tag{22}$$

and

$$\Pr\left(C^R\middle|\,\gamma_i,\xi_t,w_t\right) = \frac{\exp\left(\gamma' w_{1,t}+\xi_{1,t}\right)+\exp\left(\gamma' w_{2,t}+\xi_{2,t}\right)}{\exp\left(\gamma' w_{1,t}+\xi_{1,t}\right)+\exp\left(\gamma' w_{2,t}+\xi_{2,t}\right)+\exp\left(\gamma' w_{3,t}+\xi_{3,t}\right)}. \tag{23}$$

We note that the IIA at the individual level implies that

$$\Pr\left(d_{i1}=1\middle|\,\gamma_i,\xi_t,w_t,C^R\right) = \frac{\exp\left(\gamma_i' x_1+\xi_{1,t}\right)}{\exp\left(\gamma_i' x_1+\xi_{1,t}\right)+\exp\left(\gamma_i' x_2+\xi_{2,t}\right)}.$$

and

$$\begin{aligned} \Pr\left(d_{i1}=1\middle|\,\theta,\xi_t,w_t,C^R\right) &= \int \Pr\left(d_{i1}=1\middle|\,\gamma_i,\xi_t,w_t,C^R\right) f\left(\gamma\middle|\,\theta,\xi_t,w_t,C^R\right)d\gamma \\ &= \int \frac{\exp\left(\gamma_i' x_1+\xi_{1,t}\right)}{\exp\left(\gamma_i' x_1+\xi_{1,t}\right)+\exp\left(\gamma_i' x_2+\xi_{2,t}\right)} f\left(\gamma\middle|\,\theta,\xi_t,w_t,C^R\right)d\gamma, \end{aligned} \tag{24}$$

[6] Note that this is based on a separate contraction mapping. In the above example, the original contraction mapping was based on the 1-1 correspondence between the four-dimensional vectors. Now, the contraction mapping is a new one based on the 1-1 correspondence between the three-dimensional vectors corresponding to the first three choices.

[7] Note that there is a textbook-level identification problem, and we impose a normalization $\xi_{J,t} = 0$. For notational simplicity, we do not make the normalization explicit.

and likewise

$$\Pr\left(d_{i2}=1|\,\theta,\xi_t,w_t,C^R\right)=\int\frac{\exp\left(\gamma_i'x_2+\xi_{2,t}\right)}{\exp\left(\gamma_i'x_1+\xi_{1,t}\right)+\exp\left(\gamma_i'x_2+\xi_{2,t}\right)}f\left(\gamma|\,\theta,\xi_t,w_t,C^R\right)d\gamma,$$

where $f\left(\gamma|\,\theta,\xi_t,w_t,C^R\right)$ denotes the conditional density of γ for the subsample of individuals that chose the alternatives in C^R. By Bayes rule, we have

$$f\left(\gamma|\,\theta,\xi_t,w_t,C^R\right)=\frac{\Pr\left(C^R\middle|\,\gamma,\xi_t,w_t\right)}{\int\Pr\left(C^R|\,\gamma,\xi_t,w_t\right)f\left(\gamma|\,\theta\right)d\gamma}f\left(\gamma|\,\theta\right), \tag{25}$$

where

$$\Pr\left(C^R\middle|\,\gamma,\xi_t,w_t\right)=\frac{\exp\left(\gamma'w_{1,t}+\xi_{1,t}\right)+\exp\left(\gamma'w_{2,t}+\xi_{2,t}\right)}{\exp\left(\gamma'w_{1,t}+\xi_{1,t}\right)+\exp\left(\gamma'w_{2,t}+\xi_{2,t}\right)+\exp\left(\gamma'w_{3,t}+\xi_{3,t}\right)},$$

$$\Gamma\left(C^R\middle|\,\theta,\xi_t,w_t\right)\equiv\int\Pr\left(C^R\middle|\,\gamma,\xi_t,w_t\right)f\left(\gamma|\,\theta\right)d\gamma. \tag{26}$$

Comparison of (15) with (24) reveals a potential complication for the second step. In (15), the distribution of γ only depends on θ, whereas it depends on (θ,ξ_t,w_t) in (24). This implies that we need to fix the value of (ξ_t,w_t) <u>in addition to</u> θ when the inversion ("contraction mapping") between s_t^R and $(\xi_{1,t},\xi_{2,t})$ is performed for the subsample after the third alternative is removed. The second step in the specification test needs to address such a dual role played by the ξ's. For this purpose, we will emphasize the dual role of the ξ's and rewrite[8] (24)

$$\Pr\left(d_{i1}=1|\,\theta,\xi_t^{(1)},\xi_t^{(2)},w_t,C^R\right)=\int\frac{\exp\left(\gamma_i'x_1+\xi_{1,t}^{(2)}\right)}{\exp\left(\gamma_i'x_1+\xi_{1,t}^{(2)}\right)+\exp\left(\gamma_i'x_2+\xi_{2,t}^{(2)}\right)}f\left(\gamma|\,\theta,\xi_t^{(1)},w_t,C^R\right)d\gamma. \tag{27}$$

An intuitive idea to overcome the potential complication due to the dual role of the ξ's is to use the ξ's computed from the full set of alternatives (i.e., before removing any alternative) as $\xi_t^{(1)}$. This approach implies that the second step estimator $\widehat{\theta}_2$ may need to be based on the following complicated steps:

1. For a given candidate value of θ, use the inversion (17) for the full set of alternatives, and compute $\xi_t\left(\theta,s_t^0,w_t\right)\equiv g\left(s_t^0,w_t,F_\theta\right)$ and let $\xi_t^{(1)}=\xi_t\left(\theta,s^0,w_t\right)$ in (27).

[8] We adopted the normalization $\xi_{J,t}=0$ earlier, i.e., $\xi_{J,t}^{(1)}=0$. We note that there are only J_1 choices left after $J-J_1$ alternatives are removed. This implies that in the second step, there are J_1 such $\xi_{j,t}$'s. Therefore, the normalization should now take the form that $\xi_{J_1,t}=0$ in the second step. The two different normalization can be written $\xi_{J,t}^{(1)}=0$ and $\xi_{J_1,t}^{(2)}=0$. For notational simplicity, we do not make the normalization explicit.

2. We then view the mapping from $\left(\Pr\left(d_{i1}=1|\,\theta,\xi_t^{(1)},\xi_t^{(2)},w_t,C^R\right),\Pr\left(d_{i2}=1|\,\theta,\xi_t^{(1)},\xi_t^{(2)},w_t,C^R\right)\right)$ into s_t^R as a function in $\xi_t^{(2)}$ only, and apply the inversion ("contraction mapping") there. Note that $\xi_t^{(1)}=\xi_t\left(\theta,s^0,w_t\right)$ is a function of θ.

3. Letting

$$\tilde{g}\left(s_t^R,w_t,f\left(\cdot|\,\theta,\xi_t^{(1)},w_t,C^R\right)\right)$$

denote the result of the inversion applied to the restricted set of choices, we may then proceed with GMM adopting the moment restriction (16).

Although this idea is intuitive, it may appear to be complicated for practical implementation. We argue that the algorithm in fact simplifies quite a bit, and the specification test requires only one "contraction mapping". It turns out that in the second step, we can work with the moment equation

$$E\left[z\left(\xi_j-\xi_{J_1}\right)\right]=0,\quad \text{for all } j\in C^R$$

or

$$E\left[z\left(g_j\left(s^0,w,F_\theta\right)-g_{J_1}\left(s^0,w,F_\theta\right)\right)\right]=0,\quad j\in C^R \tag{28}$$

where ξ_{J_1} denotes the last alternative in the restricted choice set C^R. See Section 3.3 for details.

Remark 1 *We also note that the number of moment equations is smaller than when the full set of choices were considered. For example, when $J=3$ (and impose the normalization that $\xi_{t,3}=0$), the full choice set gives us two moments $E\left[z_t\xi_{t,1}\right]=0$ and $E\left[z_t\xi_{t,2}\right]=0$, whereas the restricted choice set after removing the third alternative gives us one moment $E\left[z_t\left(\xi_{t,1}-\xi_{t,2}\right)\right]=0$.*

This implies that we can use a GMM estimator that solves

$$0=\frac{1}{T}\sum_{t=1}^{T}\sum_{j\in C^R}\tilde{z}_{j,t}\left(g_j\left(s_t^0,w_t,F_{\widehat{\theta}_2}\right)-g_{J_1}\left(s_t^0,w_t,F_{\widehat{\theta}_2}\right)\right), \tag{29}$$

where $\tilde{z}_{j,t}$ denotes an arbitrary transformation of z_t, which is in general different from $z_{j,t}$ in (20).

Remark 2 *Suppose that the $z_{j,t}$ in (20) was chosen to minimize the asymptotic variance of the GMM estimator for the moment restriction (19). In other words, suppose that $\widehat{\theta}_1$ is an optimal GMM estimator. If so, we can easily see that the asymptotic variance of $\widehat{\theta}_2-\widehat{\theta}_1$ is equal to the difference of asymptotic variances of $\widehat{\theta}_1$ and $\widehat{\theta}_2$, as is usually the case with Hausman specification test. It is because (28) is implied by (18), and hence contains less information.*

3.3 Some Details behind (28)

We explain that the J_1 components of $\xi_{j,t}^{(2)}$ is equal to the first J_1 components of $\xi_{j,t}^{(1)}$ subtracted by $\xi_{J_1,t}^{(1)}$, i.e.,

$$\xi_{j,t}^{(2)} = \xi_{j,t}^{(1)} - \xi_{J_1,t}^{(1)}, \quad j = 1, \ldots, J_1. \tag{30}$$

The subtraction is just for the purpose of normalization, so we prove this property by establishing that the second contraction mapping problem can be solved by choosing $\xi_t^{(2)} = \tilde{\xi}_t^{(1)}$, where $\tilde{\xi}_t^{(1)}$ consists of the first J_1 components of $\xi_t^{(1)}$. As in the previous section, we simplify notations by assuming that $J = 3$ and that the last alternative is removed, although the analysis can be easily extended to the case with arbitrary J and J_1.

For a given value of θ, we have $\xi_{t,j} = g_j\left(s_t^0, w_t, f\left(\cdot|\,\theta\right)\right)$ in (17), i.e., the ξ's in the full sample, are computed such that if we let $\xi_{t,j} = g_j\left(s_t^0, w_t, f\left(\cdot|\,\theta\right)\right)$ in (22), it would exactly coincide with the jth component of $s_{t,j}^0$:

$$s_{t,j}^0 = P_t\left(j; \theta, g_j\left(s_t^0, w_t, f\left(\cdot|\,\theta\right)\right), w_t\right). \tag{31}$$

This implies that if we let $\xi_{t,j} = g_j\left(s_t^0, w_t, f\left(\cdot|\,\theta\right)\right)$ in (26), it would be exactly equal to the population share of C^R in the sample, i.e.,

$$\Gamma\left(C^R\middle|\,\theta, g\left(s_t^0, w_t, f\left(\cdot|\,\theta\right)\right), w_t\right) = \sum_{j \in C^R} s_{t,j}^0. \tag{32}$$

Using (24), and (25), we can write

$$\begin{aligned}
&\Pr\left(d_{i1} = 1|\,\theta, \xi_t^{(1)}, \xi_t^{(2)}, w_t, C^R\right) \\
&= \frac{\int \frac{\exp\left(\gamma_i' x_1 + \xi_{1,t}^{(2)}\right)}{\exp\left(\gamma_i' x_1 + \xi_{1,t}^{(2)}\right) + \exp\left(\gamma_i' x_2 + \xi_{2,t}^{(2)}\right)} \frac{\exp\left(\gamma' w_{1,t} + \xi_{1,t}^{(1)}\right) + \exp\left(\gamma' w_{2,t} + \xi_{2,t}^{(1)}\right)}{\exp\left(\gamma' w_{1,t} + \xi_{1,t}^{(1)}\right) + \exp\left(\gamma' w_{2,t} + \xi_{2,t}^{(1)}\right) + \exp\left(\gamma' w_{3,t} + \xi_{3,t}^{(1)}\right)} f\left(\gamma|\,\theta\right) d\gamma}{\Gamma\left(C^R\middle|\,\theta, \xi_t^{(1)}, w_t\right)}.
\end{aligned} \tag{33}$$

Letting $\xi_t^{(2)} = \tilde{\xi}_t^{(1)}$ in (33), we obtain

$$\Pr\left(d_{i1} = 1|\,\theta, \xi_t^{(1)}, \tilde{\xi}_t^{(1)}, w_t, C^R\right) = \frac{\int \frac{\exp\left(\gamma' w_{1,t} + \xi_{1,t}^{(1)}\right)}{\exp\left(\gamma' w_{1,t} + \xi_{1,t}^{(1)}\right) + \exp\left(\gamma' w_{2,t} + \xi_{2,t}^{(1)}\right) + \exp\left(\gamma' w_{3,t} + \xi_{3,t}^{(1)}\right)} f\left(\gamma|\,\theta\right) d\gamma}{\Gamma\left(C^R\middle|\,\theta, \xi_t^{(1)}, w_t\right)}. \tag{34}$$

We also note that (31) and (32) imply that

$$s_{t,1}^{0} = \int \frac{\exp\left(\gamma' w_{1,t} + \xi_{1,t}^{(1)}\right)}{\exp\left(\gamma' w_{1,t} + \xi_{1,t}^{(1)}\right) + \exp\left(\gamma' w_{2,t} + \xi_{2,t}^{(1)}\right) + \exp\left(\gamma' w_{3,t} + \xi_{3,t}^{(1)}\right)} f\left(\gamma \middle| \theta\right) d\gamma \tag{35}$$

$$s_{t,1}^{0} + s_{t,2}^{0} = \Gamma\left(C^{R} \middle| \theta, \xi_{t}^{(1)}, w_{t}\right) \tag{36}$$

Combination of (34)-(36) reveals that $\xi_t^{(2)}$

$$\Pr\left(d_{i1} = 1 \middle| \theta, \xi_t^{(1)}, \tilde{\xi}_t^{(1)}, w_t, C^R\right) = \frac{s_{t,1}^{0}}{s_{t,1}^{0} + s_{t,2}^{0}} = s_{t,1}^{R}.$$

We can similarly derive

$$\Pr\left(d_{i2} = 1 \middle| \theta, \xi_t^{(1)}, \tilde{\xi}_t^{(1)}, w_t, C^R\right) = \frac{s_{t,2}^{0}}{s_{t,1}^{0} + s_{t,2}^{0}} = s_{t,2}^{R}.$$

Because the mapping between s_t^R and $\xi_t^{(2)}$ (for given value of $\left(\theta, \xi_t^{(1)}, w_t\right)$) is one-to-one[9], and we conclude that the $\xi_t^{(2)} = \tilde{\xi}_t^{(1)}$ is the only value (up to normalization) such that

$$\Pr\left(d_{i1} = 1 \middle| \theta, \xi_t^{(1)}, \tilde{\xi}_t^{(1)}, w_t, C^R\right) = s_{t,1}^{R} \tag{37}$$

$$\Pr\left(d_{i2} = 1 \middle| \theta, \xi_t^{(1)}, \tilde{\xi}_t^{(1)}, w_t, C^R\right) = s_{t,2}^{R} \tag{38}$$

Therefore, we conclude that $\xi_t^{(2)} = \tilde{\xi}_t^{(1)}$ up to normalization. Imposing the normalization $\xi_{J_1,t}^{(2)} = 0$, we obtain (30).

4 Monte Carlo Simulations

We now present Monte Carlo simulation results of our specification tests. Our Monte Carlo design is motivated by the concern that logit specification has the well-known IIA property. See Hausman and Wise (1978), e.g., for detailed analyses of the limitations of the IIA Property as well as discussion of the alternative probit specification that overcomes the problem. See also Burda, Harding, and Hausman (2008) for further development of the alternative specification.

These simulations confirm our mixed logit specification tests have attractive size and power properties. The tests reliably identify misspecification in the mixed logit model when it exists.

[9]Note that the density of γ remains positive everywhere even after application of Bayes rule in (25), so Berry's (1994) sufficient condition for the existence of the inverse mapping is satisfied.

When there is no misspecification, type I errors occur infrequently. Below, we first describe simulations of the mixed logit specification test in section 2. Then, we describe simulation results based on the generalization of the test in section 3.

4.1 Monte Carlo Simulations of Mixed Logit Model

The Mixed Logit model described in section 2 estimates demand using individual consumers' observed choices. Our Monte Carlo simulations of this model are set up as follows. Each simulated consumer i makes choices from three choice sets that each have three options. One of the three options is an outside good for which utility is normalized to zero. Each remaining option j has an associated price (P_j) and two non-price characteristics (x_{j1} and x_{j2}).[10] We simulate choices for 500, 1000, 1500, or 2000 consumers.[11]

Consumers' simulated choices maximize their utility, which takes the same form as in equation (2) above. Preferences for non-price characteristics are drawn from a normal distribution and preferences for price are assumed equal for all simulated consumers. The first column of Table 1 below reports the parameter values used to simulate choices. For example, we assume all consumers' preferences for the first non-price characteristic are drawn from a normal distribution with a mean of 2 and a variance of 2.

Simulated choices also reflect an error term ϵ_{ij}. When we test a properly specified mixed logit model, ϵ_{ij} are drawn from an extreme value logit distribution. When we use the specification test to test a misspecified mixed logit model, ϵ_{ij} also includes an omitted characteristic that is correlated with price.[12]

Table 1 reports the parameter estimates we obtain when we estimate the mixed logit using 100 sets of simulated data, each with 2000 consumers. Column (1) reports the parameter values used to generate the data. Column (2) reports the mean maximum likelihood estimates when we estimate the original mixed logit model and there is no misspecification ($\widehat{\theta}_1$). Column (3) reports

[10] In practice, price is a randomly drawn integer between \$1 and \$10. Non-price characteristics are randomly drawn from a uniform distribution.

[11] Choice sets are allowed to vary across individuals. With 2000 consumers, for example, one simulated dataset is comprised of 6000 independently drawn choice sets (three for each consumer) that include three options each.

[12] Specifically, the omitted characteristic affecting utility takes the form $\omega_{ij} \cdot P_j$, where ω_{ij} is drawn from a uniform distribution.

the mean maximum likelihood estimates when we use the same simulated data as in column (2) but remove the outside good from the model and base estimation only on those consumers never choosing the outside option. If the model is properly specified, the estimated coefficients should be very similar under these two scenarios. Columns (4) and (5) are analogous to columns (2) and (3) except these sets of parameter estimates are based on misspecified data with endogenous prices.

The results presented in Table 1 are consistent with the intuition underlying a Hausman test. When the model is properly specified, the parameter coefficients estimated by the two versions of the mixed logit model are very similar. For example, the mean price coefficient estimated by the original model is -.499 (relative to a true coefficient of -.5). After removing the outside option from the choice set and restricting the estimation routine to those consumers never choosing the outside option, the mean estimated price coefficient is still -.499. When there is misspecification, however, the original mixed logit model generates estimates that are different than the estimates generated after removing the outside good. Under the original model the estimated price coefficient is -.184. The upward bias can be attributed to the positive correlation between price and the error term that is present under misspecification. When the outside good is removed from the model, this upward bias becomes more severe as the mean price coefficient increases to -.160. The price endogeneity differentially affects estimates of the remaining parameters as well. For example, across the 100 simulations the means of the parameters that determine the mean and variance of simulated consumers' preferences for x_1 are 1.691 and 1.498 under the full model. These estimates decrease to 1.549 and 1.096 after the outside good is removed.[13]

Table 2 reports on the size and power properties of the mixed logit specification test. The table reports results from testing two null hypotheses. First, we test the null hypothesis that all parameters (3 parameters determine mean preferences and 2 determine heterogeneity) are equal across the original model and the modified model without an outside good. Second, we test the null hypothesis that only the 3 mean parameters are equal across the two models.[14] Columns (1) and (2) report how frequently the properly specified model is rejected at the 5% level across

[13] In other words, after restricting the choice set under misspecification, the estimated preferences for x_1 change from $N(1.691, 1.498)$ to $N(1.549, 1.096)$.

[14] We include this test as an option for practitioners since it is often a challenge to precisely estimate the parameters that determine preference heterogeneity.

the 100 simulated data sets. Columns (3) and (4) report how frequently the misspecified model is rejected at the 5% level. We calculate our test statistics using three different estimates of the variance matrix. The top panel of table 2 reports test statistics that use the outer product of gradients (i.e., BHHH) as in equation (12) above.

Table 2 confirms the mixed logit specification test has desirable power properties. When the mixed logit model is properly specified, we observe type I errors infrequently. When the simulated data includes 1000 consumers and we base our test statistics on the BHHH variance matrix, we reject the null hypothesis that all of the parameter estimates (the means only) are the same across the two models in 10% (8%) of simulations. When the sample size increases to 2000 consumers, these rejection rates are 8% and 8%. When there is misspecification, the null hypothesis is frequently rejected, especially with large samples. With the BHHH variance matrix and simulated data with 1000 consumers, we reject the null hypothesis that all of the parameter estimates (the means only) are the same across the two models in 82% (89%) of simulations. With 2000 consumers, the null hypotheses are rejected in nearly 100% of simulations.

Table 2 also reports the test's power properties using alternative estimators of the covariance matrices used in the specification tests. The middle panel reports test statistics using the Hessian. The bottom panel of table 2 reports test statistics that are based on a non-parametric estimator of the variance matrix.[15] The size and power of the specification test are very similar when using BHHH or Hessians to calculate the variance matrix. The specification test has a slightly smaller size and slightly more power using non-parametric estimates of the variance matrix.

4.2 Monte Carlo Simulations of BLP

The Monte Carlo simulations of the BLP specification test described in section 3 are set up like those for the mixed logit specification test except the simulated data sets are market shares (instead of individuals' simulated choices) that reflect product level error terms (e.g., ξ_j).[16]

[15] We construct this estimate using the observed distribution of estimated parameters across the 100 simulations.

[16] Although the BLP model is estimated using aggregated market shares, the BLP specification still relies on a model of individual choice that exhibits the IIA property. Therefore, any counter-factual policy analysis based on BLP is predicated on the behavior of individual consumers who are constrained by the IIA property. Such an implicit constraint may lead to an incorrect analysis of a hypothetical merger that may result in disappearance

As before, simulated consumers make choices from choice sets that have three options. Consumers in market m in period t choose between an outside good whose utility is normalized to zero and two inside goods. Each inside option j in mt has an associated price (P_{mtj}), two non-price characteristics (x_{mtj1} and x_{mtj2}) that are observable to the econometrician and one non-price characteristic (ξ_{mtj}) that we do not observe. Products' observable characteristics (x and P) are drawn from the same distributions as described above and the product level error terms (ξ) are drawn from a normal distribution.[17]

We assume that within each market and time period, mt, all consumers face the same choice sets but we allow P and x to vary across time periods t within the same market m. While we do allow ξ to vary across time within the same market,[18] the estimation algorithm assumes that ξ_{jm} does not vary across time periods. This restriction facilitates estimation and also introduces an error into the estimation routine that is the source of variation across simulations.[19]

Within each market and time period, a continuum of consumers maximize their utility, which takes the same form as the utility function in equation (14).[20] When we apply the specification test to a properly specified model, the error terms in consumers' utility (ϵ) are drawn from an extreme value logit distribution. When we apply the specification test to a misspecified model, these errors are correlated across products within the same choice set (i.e., under misspecification, the IIA property is violated at the individual level).[21]

We simulate data for 20, 30, 40, or 50 markets, and assume that consumers within each

of certain products from the market, for example. Our specification test is developed to detect such a potential problem in the data that may distort the counter-factual analysis. In particular, our Monte Carlo design reflects the spirit of the alternative probit specification in Hausman and Wise (1978).

[17] Price (P_{mtj}) is a randomly drawn integer between \$1 and \$10. Non-price characteristics (x_{mtj1} and x_{mtj2}) are randomly drawn from a uniform distribution.

[18] Specifically, we assume $\xi_{jmt} = \xi_{jm} + \varphi_{jmt}$, where φ_{jmt} is a "shock" to product j's ξ_{jm} that varies across time.

[19] This restriction facilitates estimation because after we control for ξ_{jm} using the Berry contraction mapping, the only remaining source of variation in shares in market m across time periods is variation in X_{mt} and P_{mt}. Since the estimation routine controls for ξ_{jm} but not φ_{jmt}, draws of φ_{jmt} will affect the resulting parameter estimates.

[20] As above, preferences for non-price characteristics (x) are drawn from a normal distribution. All consumers are assumed to have the same preferences over price and ξ.

[21] Specifically, the error term that consumer i receives for product j in mt is the sum of an extreme value logit error and normally distributed error term. These latter errors are positively correlated for the outside good and the least expensive inside good.

market make choices in 10 distinct time periods. When there are 50 markets, for example, one simulated dataset includes 500 market shares (50 markets $\times$ 10 periods) for the outside good and each of the two inside goods. After simulating each dataset, we use the full choice set and the unconditional simulated markets shares $(s_{mt1}, s_{mt2}, s_{mt3})$ to estimate θ_1 using weighted nonlinear least squares.[22] We then compare these estimated parameters to the estimates we obtain (i.e.,θ_1) using a restricted choice set and the conditional market shares $\left(\frac{s_{mt2}}{1-s_{mt1}}, \frac{s_{mt3}}{1-s_{mt1}}\right)$, where s_{mt1} is the share of the outside good) and nonlinear least squares.[23] We perform this exercise 1000 times. Across the 1000 simulations, variation in the parameter estimates is driven by variation in the portion of ξ_{mj} that varies with time.

Table 3 reports the mean parameter estimates across 1000 Monte Carlo simulations when we simulate data for 50 markets and estimate BLP on properly specified and misspecified data. The results confirm that the logic of the BLP specification test is valid. When we apply BLP to properly specified data, the mean parameter estimates in columns 2 and 3 are close to the true parameter values in column 1, regardless of whether we estimate BLP using the full choice set (i.e., θ_1) or a choice set that excludes the outside good (θ_2). However, when BLP is misspecified, the parameter estimates are biased and depend on whether the model was applied to the full choice set or restricted choice set.

Table 4 confirms that the BLP mixed logit specification test has desirable power properties. First, type I errors occur infrequently. Column 1 reports the fraction of simulations when we test the null hypothesis that all of the coefficients (i.e., the mean preference parameters and the parameters that determine heterogeneity) are equal using a properly specified BLP model. The

[22] When estimating the model on the full choice set (i.e.,we estimate θ_1), we minimize the objective function $\sum_{mjt} \frac{[s_{mtj}-s_{mtj}(\theta,\xi(\theta))]^2}{w_{mtj}}$, where s_{mtj} is the simulated market share for product j in mt and $s_{mtj}(\theta, \xi(\theta))$ are the predicted market shares using θ and $\xi(\theta)$, where $\xi(\theta)$ is obtained using the Berry contraction mapping at the market level. Observations are weighted efficiently using $w_{mtj} = [s_{mtj}(\theta_{-1}, \xi(\theta_{-1})) \cdot (1 - s_{mtj}(\theta_{-1}, \xi(\theta_{-1})))]^{1/2}$.

[23] The restricted choice set excludes the outside good. So, letting $\widetilde{s}_{mtj}$ denote j's share of the inside good in mt (i.e., $\widetilde{s}_{mtj} = \frac{\widetilde{s}_{mtj}}{1-s_{mt1}}$) where s_{mt1} denotes the share of the outside good in mt), we minimize $\sum_{mt}\left[\widetilde{s}_{mt2} - \widetilde{s}_{mt2}\left(\theta, \xi_2(\theta) | \xi_3\left(\widehat{\theta}_1\right)\right)\right]^2$. Within this objective function, $\widetilde{s}_{mt2}\left(\theta, \xi_2(\theta) | \xi_3\left(\widehat{\theta}_1\right)\right)$ represents product 2's predicted share of the inside good after controlling for selection into the inside good using equation (21). We condition on the estimates of ξ for product 3 using $\widehat{\theta}_1$ because the model no longer includes an outside good. Therefore, we must normalize ξ for product 3 to the value estimated when using the full choice set.

frequency of Type 1 errors ranges from 6 percent to 8 percent depending on the number of markets we simulate data for.[24] For example, when the simulate data includes 60 markets, we reject the null hypothesis in only 8 percent of simulations. When the BLP model is misspecified, the specification test frequently rejects the null hypothesis of no misspecification. Column 3 displays these failure rates when all of the coefficients are tested. When we simulate data using only 20 markets, the misspecified BLP model fails the specification test in 59 percent of simulations. When we use 50 or 60 markets, these failure rates increase to 82 percent and 92 percent.

Table 4, we also report specification test results when we test the null hypothesis that parameters determining mean preferences do not change after re-estimating BLP after restricting the choice set (i.e., we do not test for changes in the preference heterogeneity parameters after restricting the choice set). These test results are reported in columns 2 and 4. We obtain similar failure rates as before. When BLP is properly specified and we only test the mean coefficients, the null hypothesis is rejected in fewer than 5 percent of simulations. When BLP is misspecified, we reject the null hypothesis in 54 percent of simulations that include 20 markets. This failure rate increases to 77 or 74 percent when the number of markets is increased to 50 or 60.

[24] To perform these tests, we estimate the parameters' covariance matrices non-parametrically using the distribution of estimates across the 1000 Monte Carlo simulations.

Appendix

A Our Asymptotics and Berry-Haile

We argue that Berry and Haile's (2014) identification result, if it is to be the basis of consistent estimation, implicitly requires a large-T asymptotics.

We will simplify notations by writing J instead of J_t as in Berry and Haile (2014). In their notation, the indirect utilities $(v_{i0t}, \ldots, v_{iJt})$ of agent i in market t are i.i.d. and they depend on

$$\delta_{jt} = x_{jt}^{(1)} + \xi_{jt}, \qquad j = 1, \ldots, J$$

as well as $x_t^{(2)}$ and p_t. (They drop the (1) superscript afterwards.) Individual choice is used only for the purpose of identifying the market share s_{jt} in market t. After that, the data on individual choices are not used.

Imposing primitive assumptions that justifies the contraction mapping, they obtain equation (6) on page 1760:

$$x_{jt} + \xi_{jt} = \sigma_j^{-1}\left(s_t, p_t\right).$$

They then go ahead and impose Assumption 3, which justifies z as the IV

$$E\left[\left.\xi_{jt}\right| z_t, x_t\right] = 0.$$

Assumption 4 is a completeness condition, so it is just a regularity condition. Most importantly, they identify the market share/choice probability function $\sigma_j^{-1}\left(s_t, p_t\right)$ by solving

$$E\left[\left.\sigma_j^{-1}\left(s_t, p_t\right)\right| z_t, x_t\right] = x_{jt}.$$

Solution requires working with the joint distribution of (s_t, p_t, z_t, x_t). In other words, it requires large number of t's in practice.

B s^0 vs. s^n

Consider first the infeasible estimator $\widetilde{\theta}$ which is available under the assumption that we work with the true market share s_t^0. The usual mean-value theorem applied to (20) gives us

$$\begin{aligned}\sqrt{T}\left(\widetilde{\theta}-\theta\right) &= \left(-\frac{1}{T}\sum_{t=1}^{T}\sum_{j} z_{j,t}\nabla_{\theta} g_j\left(s_t^0, w_t, F_\theta\right)\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{j} z_{j,t} g_j\left(s_t^0, w_t, F_\theta\right)\right) + o_p(1) \\ &= A^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{j} z_{j,t} g_j\left(s_t^0, w_t, F_\theta\right)\right) + o_p(1),\end{aligned}$$

which implies that $\sqrt{T}\left(\widetilde{\theta}-\theta\right)$ is asymptotically normal with the asymptotic variance equal to $A^{-1}B\left(A'\right)^{-1}$, where

$$A = -E\left[\sum_{j} z_{j,t}\nabla_{\theta} g_j\left(s_t^0, w_t, F_\theta\right)\right], \quad B = \mathrm{Var}\left(\sum_{j} z_{j,t} g_j\left(s_t^0, w_t, F_\theta\right)\right).$$

Now, consider the feasible estimator $\widehat{\theta}$ based on the estimated market share s_t^n. Assuming that

$$\frac{1}{T}\sum_{t=1}^{T}\sum_{j} z_{j,t}\left(\nabla_{\theta} g_j\left(s_t^n, w_t, F_\theta\right) - \nabla_{\theta} g_j\left(s_t^0, w_t, F_\theta\right)\right) = o_p(1), \tag{39}$$

we obtain

$$\begin{aligned}\sqrt{T}\left(\widehat{\theta}-\theta\right) &= A^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{j} z_{j,t} g_j\left(s_t^0, w_t, F_\theta\right)\right) \\ &\quad + A^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{j} z_{j,t}\left(g_j\left(s_t^n, w_t, F_\theta\right) - g_j\left(s_t^0, w_t, F_\theta\right)\right)\right) \\ &\quad + o_p(1)\end{aligned} \tag{40}$$

Therefore, as long as

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{j} z_{j,t}\left(g_j\left(s_t^n, w_t, F_\theta\right) - g_j\left(s_t^0, w_t, F_\theta\right)\right) = o_p(1), \tag{41}$$

we would obtain

$$\sqrt{T}\left(\widehat{\theta}-\theta\right) = \sqrt{T}\left(\widetilde{\theta}-\theta\right) + o_p(1) = A^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\sum_{j} z_{j,t} g_j\left(s_t^0, w_t, F_\theta\right)\right) + o_p(1),$$

i.e., the feasible estimator has the same asymptotic distribution as the infeasible one if the two high level assumptions (39) and (41) are satisfied.

We can use a textbook level discussion to develop further primitive conditions to support the two high level assumptions. For example, one can assume that (i) $\nabla_\theta g_j$ and g_j are differentiable with respect to the first argument such that the derivatives are bounded in absolute value by $\overline{G}\left(w_t, F_\theta\right)$; (ii) $E\left[\left|z_{j,t}\right| \overline{G}\left(w_t, F_\theta\right)\right] < \infty$; and (iii) $T = o\left(n\right)$, where $n = \min n_t$ and n_t denotes the number of individuals in each market t.[25] In other words, under some mild regularity conditions, we get the same asymptotic distribution as when we use the true market share s_t^0.

[25] The third assumption is useful because we have $s_t^n = s_t^0 + O_p\left(n^{-1/2}\right)$, and the second term on the right side of (40) is of order $O\left(T^{1/2}/n^{1/2}\right)$.

References

[1] Berry, S. (1994): "Estimating Discrete-Choice Models of Product Differentiation," RAND Journal of Economics 25, 242–262.

[2] Berry, S., and P. Haile (2014): "Identification in Differentiated Products Markets Using Market Level Data," Econometrica 82, 1749–1797.

[3] Berry, S., J. Levinsohn, and A. Pakes (1995): "Automobile Prices in Market Equilibrium," Econometrica 63, 841–890.

[4] Burda, M., M. Harding, and J. Hausman (2008): "A Bayesian Mixed Logit–Probit Model for Multinomial Choice," Journal of Econometrics 147, 232–246.

[5] Debreu, G (1960): "Review of R. Luce, Individual Choice Behavior," American Economic Review, 50, 186–188.

[6] Hahn, J. (1996): "A Note on Bootstrapping Generalized Method of Moments Estimators," Econometric Theory 12, 187–197.

[7] Hausman, J. (1975): "Project Independence Report: A Review of U.S. Energy Needs up to 1985," Bell Journal of Economics, 6, 517–551.

[8] Hausman, J. (1978): "Specification Tests in econometrics," Econometrica 46, 1251–1271.

[9] Hausman, J., and D. McFadden (1984): "Specification Tests for the Multinomial Logit Model," Econometrica 52, 1219–1240.

[10] Hausman, J. and D. Wise (1978): "A Conditional Probit Model for Qualitative Choice: Discrete Decisions RecognizingInterdependence and Heterogeneous Preferences," Econometrica 46, 403–426.

Table 1: Mixed Logit Parameter Estimates with and without Misspecification

	(1)	(2)	(3)	(4)	(5)
		Properly Specified Model		**Misspecified Model**	
	True Parameter	**Full Model**	**Restricted Choice Set**	**Full Model**	**Restricted Choice Set**
Preference Means					
X1	2	2.005	2.001	1.691	1.549
X2	4	4.032	4.003	3.312	3.078
Price	-0.5	-0.499	-0.499	-0.184	-0.160
Preference Heterogeneity					
X1	2	1.987	2.011	1.498	1.096
X2	3	3.079	3.204	2.104	1.491

Notes:

[1] Table reports the mean parameter estimate across 100 Monte Carlo Simulations. Each simulated dataset includes 2000 consumers who make 3 choices.

[2] The mis-specified model assumes a positive correlation between the error term and price.

Table 2: Size and Power Properties of Mixed Logit Specification Test

	(1) % of Simulations Null Hypothesis Rejected w/ Properly Specified Model	(2)	(3) % of Simulations Null Hypothesis Rejected w/ Misspecified Model	(4)
	Test All Coefficients	**Test Means Only**	**Test All Coefficients**	**Test Means Only**
Use BHHH for Variance:				
500 Consumers	12%	4%	54%	61%
1000 Consumers	10%	8%	82%	89%
1500 Consumers	11%	9%	93%	100%
2000 Consumers	8%	8%	99%	100%
Use Hessian for Variance:				
500 Consumers	13%	7%	50%	63%
1000 Consumers	9%	9%	81%	86%
1500 Consumers	10%	8%	94%	100%
2000 Consumers	10%	9%	93%	100%
Use Non-Parametric Variance:				
500 Consumers	11%	12%	84%	81%
1000 Consumers	8%	5%	100%	98%
1500 Consumers	4%	4%	100%	100%
2000 Consumers	7%	9%	100%	100%

Notes:

[1] Table reports the fraction of simulations that fail the null hypothesis of at the 5% level. Test statistics based on Hausman (1978).

Table 3: BLP Mixed Logit Parameter Estimates with and without Misspecification

	(1)	(2)	(3)	(4)	(5)
		Properly Specified Model		**Misspecified Model**	
	True Parameter	**Full Model**	**Restricted Choice Set**	**Full Model**	**Restricted Choice Set**
Preference Means					
X1	2	1.9994	1.9996	1.8738	1.8331
X2	2	1.9996	1.9993	1.8755	1.8585
Price	-0.5	-0.4999	-0.4998	-0.4689	-0.4753
Preference Heterogeneity					
X1	1	0.9993	0.9983	0.9034	1.0927
X2	1	0.9997	0.9992	0.8810	1.0061

Notes:

[1] Table reports the mean parameter estimate across 1000 Monte Carlo Simulations. Each simulated dataset includes 50 markets with a continuum of consumers who make choices in ten time periods.

[2] Under misspecification, simulated consumers' error terms includes a Logit error term and a normal error term. Draws of the normal error terms for the outside good and least expensive inside good are positively correlated.

Table 4: Size and Power Properties of BLP Mixed Logit Specification Test

	(1)	(2)	(3)	(4)
	% of Simulations Null Hypothesis Rejected w/ Properly Specified Model		**% of Simulations Null Hypothesis Rejected w/ Misspecified Model**	
	Test All Coefficients	**Test Means Only**	**Test All Coefficients**	**Test Means Only**
20 Markets	8%	4%	59%	54%
30 Markets	6%	3%	76%	62%
40 Markets	7%	1%	74%	67%
50 Markets	7%	3%	82%	77%
60 Markets	8%	3%	92%	74%

Notes:

[1] Under misspecification, simulated consumers' error terms includes a Logit error term and a normal error term. Draws of the normal error terms for the outside good and least expensive inside good are positively correlated.

[2] Null hypothesis of no misspecification tested at the 5 percent level.

[3] The simulated data includes 10 observations of each market.