Development of a Physicians' Choice Model Using Mixed Logit With Random Prices for Drugs Case Study on Diabetes Type II

Christine Claude Huttin (✉ chxyinfo@endepusresearchinc.com)
university aix marseille <https://orcid.org/0000-0002-2765-0068>

Jerry Hausman
MIT

Research article

Keywords: diabetes type II, drugs case study, physicians' choice, mixed logit

Posted Date: December 11th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-124401/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Abstract

This paper presents a first experiment with random generator of drug prices and a first simulation on physicians' treatment choices (case on pharmacotherapies) for diabetes type II care. It also aims to compare the effects of the price variables according to public versus private health plans on physicians' choices (Medicare versus commercial Health Plans). The base line model used is a Mixed Logit model with Random Price variables. A series of experiments with random parameters generations is designed with various sequences and number of draws. The model is tested on a real analytical dataset, extracted from the CDC physician survey (National Ambulatory Care Survey, NAMCS), for patients with diabetes type II without complications, for previous predictive econometrics with ENDEPUSresearch, Inc. The model uses a first drug choice set with three alternatives: oral agents only, combined therapies, no drug. The choice models introduce qualitative dependent variables and complement the series of cumulative logistic models per disease. The matlab code for the new specification test on the Independence of Irrelevant Alternatives at individual level is modified to fit this type of medical applications; first runs compare main parameters of a full choice set versus reduced choice sets of alternatives. It is planned to design more experiments for extended choice sets and widespread applications, in order to lead to user friendly tools for medical systems.

The collaboration with Professor Jerry Hausman on the US market will help with use of results and new ways to adjust the reliability on the selection of alternatives; it may provide additional guidance to the algorithms used by professionals and for health policies.

Introduction:

This paper continues the development of an economic model for medical markets, including physicians' choices models, with data on the US market; it provides a first series of results on a simulation approach, using a mixed logit model; the medical market case is the diabetic case. Data are extracted from a CDC physician survey, and models are run on an analytical database designed by Prof C Huttin and her coworkers, D Wong and S Atwood [1, 2], under her company, Endepusresearch, Inc and during visits to Harvard University (especially for analysis of ICT diffusion [3]). A new specification test to address the property of Independence of Irrelevant Alternatives (IAA) at an individual level has been developed by Prof Hausman and his co-authors [4] for mixed logit choice models and a variant of the mixed logit model (Berry, Levinsohn et Pakes (BLP) model [5]). So far, no empirical study has been performed on physicians' choices and real data. This paper provides preliminary findings on a mixed logit model on physicians' choices, with experiment on random drug prices. In this health care application, the new test may help to determine whether physicians' choices models are well specified for this dataset on Type II diabetes without complications.

Mixed logit models have already been developed in order to relax the Independent of Irrelevant Alternatives property (IAA) [6], but this property still holds at the individual level and the research proposed by Prof J Hausman and his collaborators, aims to address it, with an economic model specified

by Burda, Harding and Hausman [7]. This will overcome the limitations of nested logit models and may address in this case, all types of physicians' choices, with and without the IIA property. The new specification test [4] has first been implemented in a Matlab code by Dr J Lustig for business research and needs special modifications described in this paper for this medical application.

I. Review Of First Steps On The Development Of A Physician's Choice Model:

Prof C Huttin started this collaboration in 2017, with an invitation at MIT economic department to design physicians' choice models and transform the datasets she previously generated, under her company and research services provided by Harvard Medical school (Countway Medicine library).

The data management step is described in the first working paper from December 2018 [8]. It especially aimed to transform the selected dataset on diabetes, into a long form for qualitative choice models, to impute new random price variables, in order to shift the models from cumulative logistic or partial odd ratio models to mixed logit models. The command `asmixlogit` in Stata IC15 was used to build this economic model, the results presented at Ispor [9] help to guide the simulation approach, described in the next section of this paper, to implement the new specification test for a mixed logit model on the selected diabetic dataset. The first runs estimated parameters to obtain the convergence of the mixed logit model, using some predictors from the previous studies with cumulative logistic models [10]. The current simulation presented in this paper, implements necessary computations for this test, using both Stata license and Matlab software for the test, on the same dataset as previous models.

This diabetes care case is particularly timely in North America, with the current insulin crisis [11] and the policy proposals to import drugs from Canada [12, 13]. The development of this economic model addresses the heterogeneity of physicians' choices on a medical market, at the interaction of supply and demand. Usually, clinical choices are analyzed mostly from the supply side, within the medical system, with a specific concept of induced supply; this concept of inducement also applies on the demand side; more recent research from the principal agent literature on asymmetry of information uses the concept of induced demand. Usually business research methods are very useful to analyze these types of influence, especially with experimental design plans (e.g. conjoint models); the application to integrate patient economics with a conjoint model, called "reversed conjoint" [14, 15] clearly demonstrated it is very useful to understand how physicians' cost awareness varies with cost cognitive cues.

The economic studies on the US market by Huttin on the period 2007–2017, also validated that physicians are cost aware at the point of care (visit); for instance the 2007 study on hypertension and diabetes, provides statistical estimators on impact of patient economics on physicians' treatment choices (especially prescribing patterns), using analytical data sets extracted from the CDC National Ambulatory Care Survey (NAMCS) from CDC [16]. These statistical models (e.g. two steps models are often used in medical schools) mainly analyze the demand side of the market, especially with the impact of patients' characteristics, such as socio economic and insurance profiles and clinical and epidemiologic predictors

on physicians' choices. The statistical runs covered different periods, which allow to compare quantitative estimates of such predictors. Originally this research stream aimed to disentangle how the reimbursement design of various public/private cost sharing arrangements affect physicians at the point of care, while rules of medical ethics usually separate economic and clinical information processing. The series of predictors identified with the different models help for the statistical analysis described in this paper.

II. Description Of The Baseline Model On Mixed Logit And Modifications For The Simulation On The Diabetic Dataset:

The current model development aims to use a different approach. It is a mixed logit model with random prices for drugs and procedures. The dependent variables are physicians' qualitative choices, among various combinations/considerations of choice sets. Independent variables combine product and patient characteristics, to analyze market adjustment of supply and demand. In the first-choice experiments for the baseline mixed logit, drug prices are product attributes, they are generated with random-numbers sequences and patients' characteristics are case specific variables. This type of model is mainly coming from business research.

The base line model was implemented in a computer code, written by Dr J Lustig; the runs provide simulations of consumers' choice sets on a limited number of product characteristics. Originally, the code was written for a case scenario of 1000 consumers, three product characteristics and combinations of ten choice sets. The implementation of this test requires a comparison between combinations of three alternatives and a comparison of a reduced choice set, removing one alternative out of the three. This paper describes the modifications from the baseline model for the application on the selected medical market case: diabetes type II, without complications. This first experiment includes only one choice set: a drug choice set, while the original code could allow up to ten choice sets. This case study on physicians' choices, first needs to be described to ensure that program modifications to implement the new test, cover the various issues added to the baseline model [4]:

- The mixed logit model run with Stata, is not for consumers choices since the decision maker is the physician, who decides for or with the patient, different treatments and according to different patients' clinical and socio economic characteristics (so both product and patient' attributes)
- Physicians' choices are among alternatives on treatment choices for diabetic type II, this requires also decisions on classifications of drugs, ingredients, procedures for each alternative (usually dependent on clinical guidelines)
- This case scenario is on drug choices only, and the choice set includes three alternatives, including the "no drug" alternative; the grouping is presented in the next section with descriptive statistics of the sample
- Drug prices and the Interaction variable: Drug price for Medicare (the main federal public plan versus commercial plans) are random variables

- Three patient characteristics: age, obesity and sex are case specific variables for this simulation

Therefore, the model for this medical market case includes two random variables on products and three variables on patients. The first baseline model from Dr Lustig only includes one random variable and the case specific variables are linked to the decision maker's choices: the consumer. The random variables for this medical application are in log form contrary to the baseline model coded by Dr Lustig, who uses only normal distributions. The generation of a random variable for drug prices is also a critical step in the model development. The first working paper discusses the sources of drug price data and the selection of main common forms. This selection will have to be improved to use more complex price indexes in the model. In this paper, a generator of random numbers for prices is experimented:

The first formula for the generation of random price numbers provided by Prof J Hausman has been the following:

$$2*(0.05*Price) *uniform + price-(0.05*price) \text{ (a)}$$

Usual formula for generation of random numbers use a uniform random generation, with originally distributed random variates over an interval (0,1). But in this formula for generation of random drug prices, the random generator for prices is modified to an interval (0,2), the 5% artificial random variation is selected as a first experiment and the formula is modified into a log form of the random variable (1). So, the formula used for generation of prices for this mixed logit model is the following:

$$\text{Log} (2*runiform () + price + 1). \text{ (b)}$$

Formulas for random numbers were implemented in Stata 15C, The formula was first tested with the "asmixlogit" Stata command: "lnormal", without the log form and only with the "runiform" command for the generation of random number over the selected intervals. However, the formula with the log form and the "runiform" command in it (b) lead to more statistically significant parameters than the first formula (a) used with the lnormal Stata command. The formula avoids negative numbers in the random variates and a price equals to zero in the dataset (it helps with the "no drug" alternative) (1). Additional series of random number functions may be needed in future steps of the model development, especially to investigate correlations issues between alternatives.

Several random variables have been tested for prices since this economic model aims to adjust demand models with product prices, usually with data from the supply side. At this stage of the research, the two random prices represent general price parameters for a drug selection [9] for this type of Type II diabetic patients without complications; however the second random price variable also aims to capture the price differences paid between categories of insurance plans (these differences result to a large extent from differences in discount practices, in addition to variations in the supply chains and dispensing modes).

1. Economists usually need to use a log form in demand models of care, to incorporate very skewed distributions (long tailed), partly due to age distribution of various patient groups. Distributions of

epidemiological data, used for instance in models for disease progression, are also not normal (usual distributions are for instance Weibull, etc.).

The power sample of this dataset is sufficient to allow a comparison between public and private insurance: mainly between the federal public plan, Medicare and private plans' categories (the classification into insurance categories from the CDC was part of a detailed analysis in previous runs to estimate effects of various cost sharing categories).

Therefore, at this stage of development, the model includes two random price variables:

- one random drug price for each alternative (source: Redbook source)
- one random interaction variable: drug price x Medicare (the control group is in this case the private insurance category, other models in the future may add other categories such as Medicare advantage, Medicaid, dual eligible etc).

Mixed logit models (e.g. run with Stata software) are very used in the USA, especially since the health law passed under the Obama administration, with the creation of information exchanges on insurance plans. They are mainly used for comparative analysis of health insurance plans, using case specific variables for main types of health plans and random variables for plan characteristics such as deductibles, copayments types (e.g. tier copays) Such analysis helps to represent net price paid by patients or proxies for payment arrangements; they are often used by economists, to analyze the demand side of a medical market. However, in this paper, the random price variable in this economic model is for price adjustment on specific medical markets ; the categorical variables for insurance in the CDC survey are used to control for major differences of drug prices per main types of insurance plans, especially public versus private. This can help to examine whether there is a major difference in the coverage of patients under commercial plans and under Medicare, for age groups before 65 (especially between 55 and 65) and after 65. As this kind of modeling seems to be reliable in the statistical runs presented in this paper, it may be developed with additional types of insurance profiles and with more complex price indices for each plan category in further research. Main relevant price data providers have also been approached, a negotiation has started with Iqvia for special legacy of Pharmerit datasets (however, for the sampled case on diabetes, this dataset under-represents Medicare enrollees). Medicaid databases per state and IBM/Redbook were also consulted as possible sources. Additional sources used in main international price studies have also been reviewed and may be used for comparative analysis [17–20].

Therefore, this medical application to implement the generalization of the independence of irrelevant alternatives test and his recent specification [4] is proposed for a Mixed Logit Model including random variables for drug prices, with control variables for price differences per main type of insurance plans and patient variables, representing demographics and risk factors. At this stage, the model only includes three patient variables: age, sex and obesity. The variable age is a continuous variable, for adults over age 39 (as in previous predictive disease models already run), the cutoff point 39 is from the Diabetes clinical guidelines at the time of the study [21]. Obesity is a categorical variable in the CDC survey instrument. For

the random variables, two types of sequences have been used for the simulations: Halton and Hammersley sequences; Hammersley sequence points may also be an alternative to random numbers, they seem to significantly improve efficiency for some simulations (e.g. Monte Carlo simulation, cited in Stata Docs [22]). However, only the Halton sequence could be included at this point: a user command of Stata was needed to input the parameters of the mixed logit models, for the implementation of the new specification tests, written in a Matlab code, for the application of this code on the diabetic dataset. Stata runs with this user command, were only available with an Halton sequence by default.

iii. Results Of The Simulation With A Mixed Logit Model On The Diabetic Dataset:

The first series of mixed logit models was tested on a drug choice set. The choice set includes three alternatives of treatments in primary care, for type II diabetes without complications (the selection of the three alternatives partly results from the power sample of medical records on Type II diabetic patient):

- Alternative 1 includes the new oral agents, the oral agent continued/old.
- Alternative 2 includes combined drug therapies: insulin and oral agents, insulin and oral agent continued/old, insulin and oral new and continued/old.
- Alternative 3 is the “no drug” alternative.

The sample for this real database on physicians’ choices includes 645 medical records of diabetic type II patients (extracted from CDC NAMCS survey). Table 1 provides the descriptive statistics on the sample classified in the three alternatives A fourth alternative (injectable only) has been excluded of this analysis, with only 0.84% of the sample.

Table 1: Proportions of medical records for the three alternatives

In the sample of diabetes Type II patients

Frequencies	Alternative 1	Alternative 2	Alternative 3	Alternative excluded Injectable only
%	24,49%	39,57%	34,40%	0,84%
Source [2, 15]				

The mixed logit model was first tested with Stata command “asmixlogit” on this diabetic dataset, after transformation into a log form format. Table 2 presents some results, from this first analysis prepared for Ispor, New Orleans [8].

Table 2

```

Alternative-specific mixed logit      Number of obs   =    1,935
Case variable: id                   Number of cases  =     645

Alternative variable: alt_new        Alts per case: min =     3
                                      avg   =    3.0
                                      max   =     3

Integration sequence:  Hammersley
Integration points:    50
Log simulated likelihood = -681.4805      Wald chi2(8)    =    15.54
                                      Prob > chi2       =    0.0494

```

choice_new	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
alt_new						
price_new	-.4377095	.5272314	-0.83	0.406	-1.471064	.595645
paymcare#c.price_new 1	.1453872	.1353147	1.07	0.283	-.1198247	.4105992
Normal						
sd(price_new)	.4411865	.425822			.0665371	2.925367
sd(0.paymcare#price_new)	0	(omitted)				
sd(1.paymcare#price_new)	.9485012	.5462338			.3067878	2.932498
AT_LEAST_1_NEWDRUG						
age	.0331875	.0100668	3.30	0.001	.013457	.0529181
female	-.4793454	.2242222	-2.14	0.033	-.9188128	-.0398779
obesity	-.1954772	.2799411	-0.70	0.485	-.7441517	.3531973
_cons	-2.409955	.8760187	-2.75	0.006	-4.12692	-.6929898
BOTH	(base alternative)					
NEITHER						
age	.0231909	.0154547	1.50	0.133	-.0070997	.0534815
female	.111055	.3735315	0.30	0.766	-.6210532	.8431632
obesity	-.4687868	.4970349	-0.94	0.346	-1.442957	.5053836
_cons	-4.188285	3.095195	-1.35	0.176	-10.25476	1.878187

LR test vs. fixed parameters: chi2(2) = 4.70 Prob > chi2 = 0.0955

Note: LR test is conservative and provided only for reference.

The parameters on physicians' choices, estimated on the real diabetic dataset, with the mixed logit model presented in Table 1, are used in the simulation approach presented in this paper; the simulation is implemented in matlab code. In order to use the new specification test proposed by Professors Hahn, Hausman and Lustig (2020) [4], the original code programmed by Dr Lustig had to be modified for the purpose of this specific medical application. The Independence of Irrelevant Alternatives (IIA) property can then be examined for physicians' choices, at an individual level, for the drug choice set including Alternatives 1, 2 and 3. The test compares estimated parameters between the choice set of three alternatives, including the "no" diabetic drug, and a reduced choice set of two alternatives out of three; the removed alternative is either the first alternative, the second or the third one; Normally, if the specification test is verified, coefficients of the parameters between the choice set of three versus a reduced choice set of two alternatives, will be quite similar, whatever the alternative which is removed from the reduced choice set.

The statistical runs using a mixed logit model on the drug choice set are simulations with Halton sequences of 50 draws or 200 draws on the real dataset of diabetes type II. The results provided in this paper are only on the 200 draws simulations, since the Log likelihood Ratio is significant (Prob > Chi2 = 0.05) only with runs with 200 draws, and not with the default number of 50 draws.

Tables 3 and 4 provide some examples of results of the simulations: table 3 gives physicians' preferences variations with prices, when alternative 1 ("on oral agent only") is removed and Table 4 when alternative 2 ("on combined therapy") is removed. They show the comparisons of coefficients for some parameters, between the full choice set and the reduce choice set.

Table 3: Mixed logit parameter estimates (for specification test)

(Simulation of 200 draws)

Mean preferences	Full Model with three alternatives (Alternatives 1,2,3)	Restricted Choice set (Alternatives 2 and 3)
Price	- 2.4270	2.8546
Price x Medicare	0.4438	0.4227

Table 4: Mixed logit parameter estimates (for specification test)

(Simulation of 200 draws)

Examples of some comparisons on price (random) and age (case specific variable)

Mean preferences	Full Model with three alternatives (Alternatives 1,2,3)	Restricted Choice set (Alternatives 1 and 3)
Price	- 2.4270	-1,4016
Price x Medicare	0.4438	-0.8990
Age x alternative 1	-7,0235	-8,9349
Age x alternative 3	-20,1947	-21,3823

Tables 3 and 4 show the comparisons of coefficients between full and reduce choice sets in two cases out of three (they will be used for the application of the new specification test on IIA [4]):

- when alternative 1 is removed for the reduced choice set, the two coefficients associated with the random drug price for Medicare are quite similar between the two choice sets as well as the two coefficients associated with the interaction variable between obesity and alternative 2.
- When alternative 2 is removed, it is the main random price variable (and not the one associated with Medicare) which shows negative coefficients in both choice sets, all coefficients associated with the interaction variables between age and alternative 1 (one oral agent) and age and alternative 3 (no

drug) are similar for the choice set with three alternatives and the reduced choice set of 2 alternatives.

- When alternative three (no drug case) is removed, the comparison of the two maximum likelihood functions does not work at this point; additional modeling steps are investigated; in particular since the reduced choice set in this case includes the two alternative strategies with pharmacotherapies, different generators of random prices may be needed for the two different drug therapies, in order to avoid correlation issues (e.g. in sequences).

IV. Discussion And Limitation:

The computer simulation tends to compare a choice set with three alternatives (mainly two alternatives of a drug choice (either at least one oral agent or a drug choice of a combined therapy), versus a no drug choice. When a drug alternative with an uptake of drugs “either at least one oral agent “ (Alt 1) or a drug alternative “combined therapy” (Alt 2) is removed for the reduced choice set, the test specification at individual level tends to show that there is independence of such alternatives, since some coefficients of some parameters are quite similar. However, when the alternative “ no drug” is removed from the reduced choice set, the specification test comparing a reduced choice set with two drug therapies with the choice set of three alternatives (which included the no drug alternative) does not work.

Can we discuss at this point, the independence of irrelevant alternatives property at individual level of physicians? In the pre diabetic stage, it may not be relevant to be on a drug therapy (e.g. controversy between pharmacists and physicians at the time of the study on prescribing an oral agent versus aggressive preventive strategy). It may explain that when the “no drug alternative” is removed, the new test comparing a reduced choice set comparing two different clinical strategies with drug treatment and a choice with no drug, reflect irrelevant alternatives, for some individuals. However, further statistical analysis is needed to explore potential estimation issues for implementing the new test for all alternatives on this diabetic dataset.

V. Conclusion:

This type of simulation is often used in the field of preference research especially in health care for patient preferences, using Discrete Choice Analysis (DCE) or shared decision making decision models; current methodological papers [23] discussed the benefits of simulated multinomial logit according to number of draws, investigating the effects of correlation between Halton sequences on multinomial estimates according to number of draws. The IHAPR academy and the transportation research econometric society recommend performing sensitivity analysis according to the number of random parameters and the number of draws (to revise and complete Chikososky, 2016; finance literature); the two runs on 50 and 200 draws already show the changes on the likelihood ratios when the number of draws is increased; however, additional sensitivity analysis may be necessary to examine the effects of correlation between sequences. As in the first study using “asmixlogit” Stata command (Table 1), other

sequences than Halton sequences may also improve the results. The runs with an Hammersley sequence, seemed to provide more reliability of some parameters.

Abbreviations:

IIA: Independent of Irrelevant Alternatives

IAHPR: International Academy of Health Preference Research

Declarations:

Acknowledgement:

the authors are grateful to the Data Science center from Harvard-IQSS for their support on the project and to Dr Lustig for his role in the first code.

Ethics approval and consent to participate

No need for approval at this stage

This type of research does not require IRB approval Or other ethical approval

The researchers have used the official codes of the CDC at the time of the survey Additional data input are also coming for public sources (red book and its online version for the prices)

Other sources of data may be used for larger applications

Drug Choices are selected following the consensus on the clinical guidelines at the time of the survey

Consent for publication

Prof Chritine C Huttin agrees to publish this paper

Prof J Hausman

Availability of data and material

Data are extracted from the CDC Physician survey, this administrative dataset is publicly available from the agency for the health service research community

The extraction methodology is described in several papers listed in the reference lists and could be requested to authors

The price information used are also publicly available from the Redbook price lists , collected in medical library and also available online

Competing interests

Funding

This is a non funded research project

The no funding letter has been added for the waiver fee resolution by Prof J Hausman from MIT , the institution that contributes to the APC

Authors'contributions

Need to check what is requested?

References:

1. Huttin C and Wong R "Comparative results on effects of type of insurance and payment plans on physicians' cost awareness during the physician visit, using the NAMCS survey in 1996 and 2005, Value in Health, June 2010.
2. Huttin C and Atwood S IT process in clinical practices for diabetes type II. Value in Health, May 2011:14(3): A103.
3. Huttin C "IT diffusion and Primary care methodological contribution on clustering methods to partition medical practices in the USA" Int Journal of Computer Science and Mobile Computing (IJCSMC), Vol 9, Issue 4, April 2020:49-60.
4. Hahn J, Hausman J, Lustig J "Specification test on Mixed Logit Models" Journal of Econometrics, Jan 2020.
5. Berry S, Levinsohn J, Pakes A Automobiles prices in market equilibrium Econometrica, Vol 63, N4, 1995:841-890.
6. Hausman J and McFadden D; Specification tests for the multinomial logit model, Econometrica, Vol 52, N5,121961240, Sept 1984.
7. Burda M, Harding M, Hausman J "a Bayesian Mixed Logit-Probit Model for Multinomial choice" Journal of Econometrics, 147,232-246,2008.
8. Huttin C, Hausman J "Development of a choice model for physicians with random preferences applications on diabetes type II", Value in Health, Vol 22, Issue S1, May 2019.
9. Huttin C and Hausman J Working paper N. Physicians' choice models on a diabetes dataset, unpublished-Dec 2018, MIT.
10. Huttin C "Stated and Preference issues: Product, Patient and Physicians' attributes in choice experiments" Technology and Health Care 25(2017) 1005-1020, <https://doi.10.3233/THC 160769>
11. Fralick M, Avorn J, Kesselheim A The price of crossing the border for medications. N Engl J Med 377;4 2017:311:313. DOI:10.1056/NEJMc1711278
12. Ashley D The price of crossing the border for medications, N Eng J Med 377;17; Oct 2017: 1699.

13. Luo j, Avorn J, Kesselheim A Trends in Medicaid reimbursements for insulin from 1991 through 2014. *Jama intern Med* 2015;175 (10):1681-1687.
14. Huttin C New type of price measurement for medical services; interest in a cost sensitivity index for a research agenda on pharmaceutical models; *Technology and Health Care* 20 (2012): 463-476.
15. Huttin C Economics and medical decision making, series of methodological papers Ed Scholars Press, May 2017
16. Huttin C “The role of different types of health insurance on access and utilization, comparative results on two chronic conditions” *Managerial economics*, Wiley, 28:503-5-7, 2007.
17. Redbook online service to complete
18. Medicaid price: www.data.medicaid.gov
19. OECD, pharmaceutical pricing policies in a global market, 2008.
20. American Diabetes Association guidelines ref to complete
21. Simmons Stata Docs, 2005
22. Ellis A, de Bekker-Grob E , Howard K, Thomas K, Lancsar E, Ryan M, Rose J “ Number of Halton Draws Required for Valid Random Parameter Estimation with Discrete Choice Data” *Proceedings, 10th Meeting of the International Academy of Health Preference Research (IAHPR)*

Other useful references

Burda M, Harding M, Hausman J, “A Poisson Mixture Model of Discrete Choice”, *Journal of Econometrics*, 166,2012.

Baht C Quasi random maximum likelihood simulated likelihood estimation of the mixed multinomial logit model 3 working paper department of civil engineering university of Texas Austin, 1999.

Huttin C “Choices of multi-stakeholders for economic models towards precision medicine” Abstract S201701214, *Proceedings, EUPHA-10th EPH conference, Stockholm, European Journal of Public Health, Supplement, Nov 2017.*

Huttin C “Modeling of demand for care and health insurance, appropriate level of aggregation, Chapter IV, in *Economics and Medical Decision Making*, ed Scholars’ Press, 2017.

McFadden D and Train K Mixed MNL models for Discrete Response, *Journal of Applied econometrics*, 15:447-470, 2000.

Train K Halton sequences for Mixed logit <https://escholarship.org/UC/item/6zs694tp> and additional references on <http://elsa.berkeley.edu/train>

Appendix:

Appendix 1: summary of the mathematics under the simulation of mixed logit

“Choice probabilities under mixed logit take the form of a multinomial integral over a mixing distribution” (Brownstone and Train, 1999). The integral is therefore evaluated numerically with the simulation draws.

The formulation of the mixed logit is the following:

If Lin is the logit formula

$P_{in} = \int Lin(b) f(b) db$

The integral of the mixing distribution is estimated with either random draws or draws from different distribution (usually Halton draws).

Appendix 2 sequences and simulation in mixed logit

The simulation of parameters on mixed logit models have usually been performed with Halton sequences instead of random draws since the early studies by Bhat, 1999; other findings especially at UC Berkeley with Train, (e.g. 2000) confirmed the improvements of parameters estimates with such sequences (1); and usually software include this type of sequences for simulated results (e.g. MATLAB or Stata). Moreover, research in this field showed the need for a large number of draws in order to have reliable estimates of the parameters.

1. For interested readers, Baht’s comparisons of simulation on mixed logit using Halton sequences versus random sequences provide statistically significant improvement on estimates of results (to complete)

Appendix 3: Explanation on the statistical results on price random parameters

These statistical results compare results for a drug choice set with three alternatives and a reduced choice set, for computation of the generalization of the Hausman and McFadden test on the diabetic type II dataset.

The determination of the alternative to be removed is usually called the statu quo or outside good. In this application, the “no drug” alternative was first considered as the one to remove for the reduced choice set, with inclusion of the two different drug therapies in the reduced choice set: however. This is not obvious on diabetic type II (see controversies on classification and preventive strategies); therefore, each alternative was successively removed: alternative 1, alternative 2 and alternative 3 for the reduced choice sets. Then, three comparisons between simulated parameters for three alternatives versus two alternatives were run, using 50 and 200 draws. The simulations used a different version of the mixed logit model; it was run a second time, using the same predictors on the dataset, but with a Harvard Stata in house license, in order to transfer coefficients from the Mixed logit model into MATLAB code. Tables 3 and 4 present preliminary comparisons of coefficients, for the simulated random price parameters and age. This code is currently under revision and needs additional modifications for various applications.