Problem Set Simulated Data Description

Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts

Ben Handel

This document describes the simulated dataset 'ASIN-ChoiceModelData-FINAL.mat' provided along with the problem set relating to the paper *Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts* by Ben Handel. These data are 'processed' or set up to be a form that is close to the form necessary to estimate the primary choice model described in the paper and discussed in the problem set.

The following table describes the list of simulated variables and their dimensionality in Matlab. Note that while some parts of the paper use individual level data (e.g. the cost model) the observations in the choice model estimation data below are at the family level since this is the relevant unit of choice.

Variable Name	Dimensionality	Description
nIs	Scalar	Number of families in data, equal to 2290 in simulated data to mimic actual choice model estimation
K	Scalar	Number of simulated ex post health draws to model distribution of ex ante health risk, set = 50
Sim	Scalar	Number of simulation draws to describe distributions for risk preference coefficients, epsilons, and PPO1200 / CDHP random coefficients. Set = 50 here.
nPlans	Scalar	Number of health plans being offered / chosen between, = 3.
nTs	Scalar	Number of periods in panel that consumers make choices. Here, set = 3 to mimic actual environment.
P1	(nTs,nIs,Sim)	Prices for PPO250 for employees over three years, filled in for each risk preference distribution draw. Price varies by employee with income / family status.
P2	(nTs,nIs,Sim)	Prices for PPO500 for employees over three years, filled in for each risk preference distribution draw. Price varies by employee with income / family status.

P3	(nTs,nIs,Sim)	Prices for PPO1200 for employees over three years, filled in for each risk preference distribution draw. Price varies by employee with income / family status.
choice	(nIs,nTs)	Health plan choice made by each family in each year. 1=PPO250, 2=PPO500, 3=PPO1200. Choices are simulated in prior file with preferences similar to those found in actual estimation. Code for simulating choices is available upon request.
Ages[1]	(nIs,4)	Matrix of ages for people in each family. Maximum number of people in family in simulated data is 4, which is why that is second dimension.
CC1 & CC2	(nIs,1)	Variable indicating whether individual / family have chronic conditions present in year 1 or year 2 of data. For individual CC = 1 if chronic conditions > 1, for family if chronic conditions > 2.
CSAL2 & CSAL3	(nIs,1)	Indicator for large change in past year, =1 for year 2, if, in year 1, family expenditure changed by >50% (either up or down) relative to previous year expenditure (year 0).
FSAY1 & FSAY2 & FSAY3	(nIs,1)	Indicator whether family is enrolled / contributing to a flexible spending account (FSA) in either year 1,2, or 3.
FamidX	(nIs,1)	Family random ID number in case necessary for code
Famsize	(nIs,1)	Variable describing number of members in each family
Genders	(nIs,4)	Matrix with genders for each member in each family. Second dimension equals 4 because

---

[1] NOTE: The matrix Ages, Genders, and Famsize DO NOT match up correctly with the family status variables Tier2 and IND due to the way the simulated data were created. Ages and Genders do reflect the number of people used to simulate each family, but the IND and Tier2 variables were created orthogonally as 'preference' variables only to use in choice model. The problem set describes how to appropriately use these variables in estimation in a straightforward manner.

		maximum number of members in family in simulated data is 4.
HTCi	(nIs,1)	High total cost indicator used in estimation of preferences. In simulated data, created as families that have greater than $27,000 in expected expenses for year 2. Meant to capture high expenses generally, and used as constant over panel.
IND	(1,nIs)	Vector indicating whether family in question is comprised of one individual only (single, =1) or more than one person (family, =0)
Inc1 & Inc2	(nIs,1)	Income variables for families in years 1 and 2. Income is grouped into 5 tiers as in actual data. In actual data, Tier 1 is aprox. 0-$40000, Tier 2 $40,000-$80,000, Tier 3 $80,000-$120,000, Tier 4 $120,000 to $160,000, and Tier 5 $160,000 and above.
PPO250OOP1 & PPO250OOP2 & PPO250OOP3	(nIs,K)	Matrix describing K out-of-pocket expenditure draws for each family and for PPO250, given their ex ante health risk and plan characteristics. These variables are derived in detailed in cost model estimation code supplied. Provided for each year 1,2,and 3.
PPO500OOP1 & PPO500OOP2 & PPO500OOP3	(nIs,K)	Matrix describing K out-of-pocket expenditure draws for each family and for PPO500, given their ex ante health risk and plan characteristics. These variables are derived in detailed in cost model estimation code supplied. Provided for each year 1,2,and 3.
PPO250OOP1 & PPO250OOP2 & PPO250OOP3	(nIs,K)	Matrix describing K out-of-pocket expenditure draws for each family and for PPO1200, given their ex ante health risk and plan characteristics. These variables are derived in detailed in cost model estimation code supplied. Provided for each year 1,2,and 3.
PlanPaid1 & PlanPaid2 &	(nIs,K,nPlans)	Matrix describing how much

PlanPaid3		insurance plan pays for each family, for each health draw, for each plan. 1,2,3 correspond to years. PlanPaid here equals total cost minus OOP.
QS	(nIs,1)	Indicator for whether person is employed in "quantitatively sophisticated" job. In actual data, this was done with categorization of job data that cannot be made public, but there are clear delineations within firm.
Tier0 & Tier1 & Tier2	(nIs,1)	Variable giving family status for each family in each year 0,1,and 2. Tier2 is also applied to year 3 as in actual data (since we observe year 3 choices but not demographics). 1=single, 2 = +spouse, 6=+child(ren) and 8=+spouse+child(ren). Plan premiums depend on family status and we condition some estimates on this estimation.
Total1 & Total2 & Total3	(nIs,K)	Total medical expenditures for family + insurer for each health draw and each year (1,2,or3). Under no moral hazard assumption, this is same across plans.
managerX	(nIs,1)	Variable indicating whether employee was 'manager' or high-level employee at the firm. Linked to inertia in choice model and constant over time assumed.