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

Benjamin R. Handel

Data Description

This document describes in detail the two simulated datasets provided to the *American Economic Review* for distribution for "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts." As noted in the README file, also included for distribution, the author can only provide simulated data because the actual data used in this study (i) are proprietary (no release agreement with data provider) and (ii) are protected from public release by HIPAA / IRB under the regulations governing a 'Limited Data Set.' Please contact the author at handel@berkeley.edu with further questions on this.

There are two simulated data files generated to use with the code provided with this submission. The primary file provided is a simulated data file for choice model estimation, where cost model estimation and processing has already been incorporated from prior simulated data / code. After describing this primary file, we describe the interim simulated data used for the cost model processing code provided (a step which precedes choice model estimation).

The primary simulated data for choice model estimation is in file 'ASIN-ChoiceModelData-FINAL_2011_1284.mat.' These data are called by the code 'EstimationCode.m' to perform choice model estimation. The following table describes the list of simulated variables and their dimensionality in MATLAB. NOTE: everything in the choice model estimation data below is in FAMILY form. Cost model data described afterwards is at INDIVIDUAL level (and later aggregated to family for choice model estimation).

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	(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 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 & PlanPaid3	(nIs,K,nPlans)	Matrix describing how much 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.

The other simulated data included with the submission are an interim data set that is designed to be used with cost model simulation to prepare health risk distributions that are ultimately used in choice model estimation. These data are in the file "ASIN-Handel-SimulatedData_2011_1284.mat." They are used as inputs into the cost model simulation / implementation code provided in the file "Cost_Model_Implementation_2011_1284.m." This code processes the data to be used in choice model. Note: output of this code cannot be used an input into choice model estimation here, because there is interim step to simulate choices in simulated data. For choice model estimation use files described above which incorporate simulated choices. This material is provided to show how to run cost model implementation with actual data where choices are present. The following table describes the list of simulated variables and their dimensionality in MATLAB I the file "ASIN-Handel-SimulatedData_2011_1284.mat."

Variable Name	Dimensionality	Description
nIs	(nIs,1)	(NOTE: Defined in first line of cost model implementation code). Number of individuals in data. NOTE: nIs here represents individuals. In choice model data described above it represents number of families, or aggregated groups of individuals so it's a smaller number in other data.
ChronCt1 & ChronCt2	(nIs,1)	Number of chronic conditions identified in years 1 & 2 by ACG program in actual data (here simulated).
CorrOVH1 & CorrOVH2 & CorrOVH3	(nIs,1)	Spearman rank correlation estimated in cost model estimation code. Rank correlation in ex post realizations in Office Visit expenses and Hospital / Other expenses. See cost model estimation for more details: this is correlation in projected expenditures in upcoming year.
CorrOVRX1 & CorrOVRX2 & CorrOVRX3	(nIs,1)	Spearman rank correlation estimated in cost model estimation code. Rank correlation in ex post realizations in Office Visit expenses and Pharmacy expenses. See cost model estimation for more details: this is correlation in projected expenditures in upcoming year.
CorrRXH1 & CorrRXH2 & CorrRXH3	(nIs,1)	Spearman rank correlation estimated in cost model estimation code. Rank correlation in ex post realizations in Pharmacy expenses and Hospital / Other expenses. See cost model estimation for more details: this is correlation in projected expenditures in upcoming year.
FSA1 & FSA2 & FSA3	(nIs,1)	Indicator of whether individual belongs to family that enrolls in flexible spending account in each of three years.
HTier0 & HTier1 & HTier2	(nIs,1)	Variable giving family status for

		each family (that individual in question belongs to) 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.
Hb1 & Hb2 & Hb3	(nIs,1)	Weibull shape parameter for Hospital / Other expenses for each individual in each year for projected expenses in that category.
Hc1 & Hc2 & Hc3	(nIs,1)	Weibull scale parameter for Hospital / Other expenses for each individual in each year for projected expenses in that category.
Individual	(nIs,1)	Indicator whether individual belongs to a family or not. =1 if single.
MHAgeCoeffb1 & MHAgeCoeffb2 & MHAgeCoeffb3	(nIs,1)	Weibull shape parameter shifter as a function of individual age. See cost model code for further details. For mental health category expenditures here, projected for upcoming year.
MHAgeCoeffc1 & MHAgeCoeffc2 & MHAgeCoeffc3	(nIs,1)	Weibull scale parameter shifter as a function of individual age. See cost model code for further details. For mental health category expenditures here, projected for upcoming year.
MHGenderCoeffb1 & MHGenderCoeffb2 & MHGenderCoeffb3	(nIs,1)	Weibull shape parameter shifter as a function of individual gender. See cost model code for further details. For mental health category expenditures here, projected for upcoming year. NOTE: gender=1 if male.
MHGenderCoeffc1 & MHGenderCoeffc2 & MHGenderCoeffc3	(nIs,1)	Weibull scale parameter shifter as a function of individual gender. See cost model code for further details. For mental health category expenditures here, projected for upcoming year.

MHb1 & MHb2 & MHb3	(nIs,1)	Weibull shape parameter intercept for Mental Health expenses for each individual in each year for projected expenses in that category.
MHc1 & MHc2 & MHc3	(nIs,1)	Weibull scale parameter intercept for Mental Health expenses for each individual in each year for projected expenses in that category.
OVb1 & OVb2 & OVb3	(nIs,1)	Weibull shape parameter for Office Visit expenses for each individual in each year for projected expenses in that category.
OVc1 & OVc2 & OVc3	(nIs,1)	Weibull scale parameter for Office Visit expenses for each individual in each year for projected expenses in that category.
PrHHP1 & PrHHP2 & PrHHP3	(nIs,1)	Probability estimated in cost model of having hospital / other expenditures > 40,000. This is estimated as discrete probability separate from Weibull estimates for this category so estimates aren't affected a lot by large tail expenditures here. See cost model estimation code for more details.
PrZHP1 & PrZHP2 & PrZHP3	(nIs,1)	Part of two part distribution described in cost model estimation describing discrete mass at 0 expenditures. Distribution has mass at 0 discrete then positive expenditures follow Weibull. This is for hospital / other expenditures.
PrZMH1 & PrZMH2 & PrZMH3	(nIs,1)	Part of two part distribution described in cost model estimation describing discrete mass at 0 expenditures. Distribution has mass at 0 discrete then positive expenditures follow Weibull. This is for mental health expenditures.
PrZRX1 & PrZRX2 & PrZRX3	(nIs,1)	Part of two part distribution described in cost model estimation describing discrete mass at 0 expenditures. Distribution has mass at 0 discrete then positive expenditures follow Weibull. This is

		for RX expenditures.
PrZOV1 & PrZOV2 & PrZOV3	(nIs,1)	Part of two part distribution described in cost model estimation describing discrete mass at 0 expenditures. Distribution has mass at 0 discrete then positive expenditures follow Weibull. This is for office visit expenditures.
QuantSoph	(nIs,1)	Indicator whether individual belongs to family where employee is classified as working in a 'quantitatively sophisticated' job. Linked to inertia in choice model estimation.
RXAgeCoeffb1 & RXAgeCoeffb2 & RXAgeCoeffb3	(nIs,1)	Weibull shape parameter shifter as a function of individual age. See cost model code for further details. For pharmacy category expenditures here, projected for upcoming year.
RXAgeCoeffc1 & RXAgeCoeffc2 & RXAgeCoeffc3	(nIs,1)	Weibull scale parameter shifter as a function of individual age. See cost model code for further details. For pharmacy category expenditures here, projected for upcoming year.
RXGenderCoeffb1 & RXGenderCoeffb2 & RXGenderCoeffb3	(nIs,1)	Weibull shape parameter shifter as a function of individual gender. See cost model code for further details. For pharmacy category expenditures here, projected for upcoming year. NOTE: gender=1 if male.
RXGenderCoeffc1 & RXGenderCoeffc2 & RXGenderCoeffc3	(nIs,1)	Weibull scale parameter shifter as a function of individual gender. See cost model code for further details. For pharmacy category expenditures here, projected for upcoming year.
RXb1 & RXb2 & RXb3	(nIs,1)	Weibull shape parameter intercept for Pharmacy expenses for each individual in each year for projected expenses in that category.
RXc1 & RXc2 & RXc3	(nIs,1)	Weibull scale parameter intercept for Pharmacy expenses for each individual in each year for projected expenses in that category.
age	(nIs,1)	Vector of ages for individuals in the

		sample.
famid	(nIs,1)	Random family ID number that associates each individual with a family.
famsize	(nIs,1)	Vector of family sizes for families associated with each individual.
gender	(nIs,1)	Vector of genders associated with each individual. 1=male
income1 & income2	(nIs,1)	Income for each family individual is associated with in each year 1 & 2
manager	(nIs,1)	Indicator whether employee in family is high level employee or 'manager' within the firm.