

Read-me for data file simulateddrugs01-sas7bdat

Farasat A.S. Bokhari

f.bokhari@uea.ac.uk

Centre for Competition Policy, University of East Anglia, Norwich NR4 7TJ, UK

School of Economics, University of East Anglia, Norwich NR4 7TJ, UK

November 10, 2014

Introduction. This document provides the structure of, as well as the list of variables, in the SAS datafile ‘simulateddrugs01-sas7bdat’. As the name suggests, the data is simulated: it consists of simulated sales data for 17 ADHD drugs and should not be used for drawing any conclusions about the ADHD drugs market. It is, however, based on the data described in [Bokhari and Fournier \(2013\)](#) and the simulated values are derived from an estimated model described in the paper. The data is set in ‘wide’ format meaning it has 3112 observations from 778 counties (from 48 states and Washington D.C.) and upto four years (2000-2003). Thus, a county-year is a market and is one of the 3112 rows, while sales data (quantities, revenues, prices, etc.) and some demographic information for each of these markets is given in the columns. Further, the data is not balanced in the sense that not all drugs were on the market for the full four years and hence the simulated data maintains the new products problem where sales for a product are obviously not observed prior to its marketing.

Area Identifiers and Demographics (Variables 1-25). The first 25 columns identify the county (fips code), state, census division, census region, year and also provide some basic demographic information about these counties such as population, number of children, etc. in each county (since this information is public, real values of these demographic variables is included). The variables names and labels are as given in [Table 1](#) below.

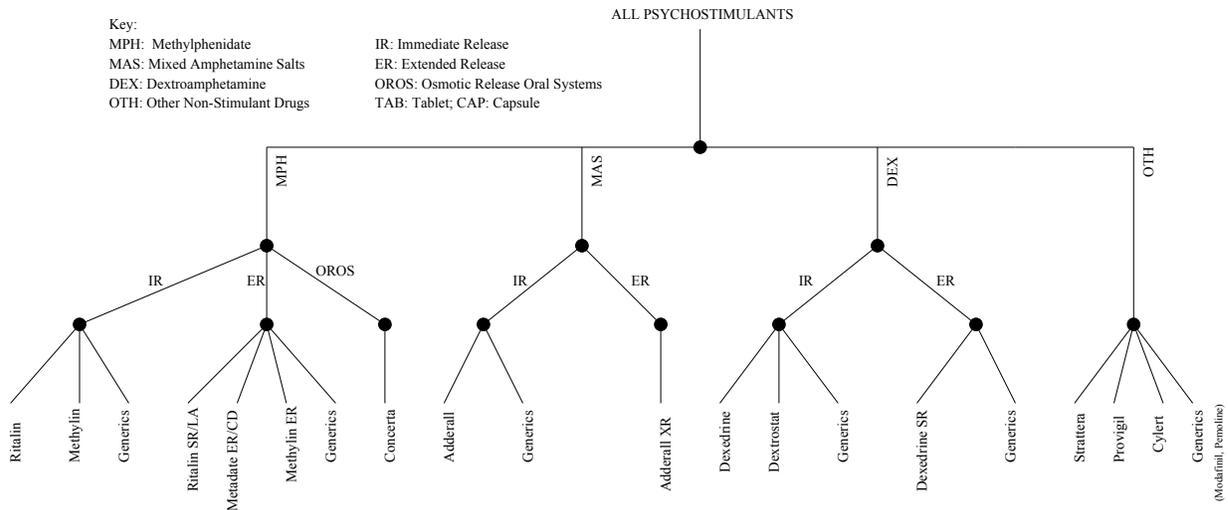
#	Variable	Label
1	fips	5Digit Fips State-County Code, Character
2	stabr	State Name Abbreviation
3	year	Year
4	msac	4Digit MSA/PMSA Code, Character
5	cntyst	County Name W/State Abbrev
6	cenreg	Census Region Name
7	cenregc	Census Region Code
8	cendivn	Census Division Name
9	cendivc	Census Division Code
10	msapmsa99c	MSA/PMSA Code 1999
11	poptot	Total Population in 1000s
12	kids0t4	Children (Age 0 to 4)
13	kids5t19	Children (Age 5 to 19)
14	kids	Children Aged 5-19
15	pcpi	Per Capita Personal Income (Real 2000 \$s)/1000)

#	Variable	Label
16	mds	Total MDs in 1000s (MDTPC)
17	caiddrugs	Amount Reimbured for Drugs by Medicaid
18	mcaidenrollees	Medicaid Enrollees in State (Census Estimates)
19	lncaiddrugs	Log (Total Amount Reimbured for Drugs by Medicaid)
20	lnmcaidenrollees	Log(Medicaid Enrollees in State (Census Estimates))
21	lnpop	
22	lnkids	
23	lnmds	
24	t1	Time (t1): 1 is year 1999
25	t2	Time Sq (t2): 1 is year 1999

TABLE 1. Area/year identifiers and demographic variables

Revenues by group of drugs and total (Variables 26-38). While there are 17 basic drugs, they are grouped in a 4 level system: at the bottom level are individual drugs that are in the same molecule and form (and hence there are 8 such bottom level groups), next level up, drugs are aggregated into forms within the molecule (and hence there are four such groups), further up is molecules within all drugs and then all drugs. See [Figure 1](#).

FIGURE 1. Taxonomy of ADHD drugs by Molecule, Form and Brand Names



Note: Generics refer to several manufacturers for each molecule and form given in the column. There are no generic versions of Concerta and Adderall XR during the study period.

The revenues at different level of aggregation are given in variables 26-38: variable **rr** lists the total revenue (or expenditure) on all drugs while **y1** through **y8** provide revenues at the bottom level. For instance **y1** is the total sales revenue from three drugs – Ritalin, Methylin and Generic (drugs 1,2,3) which are in the MPH-IR group. Similarly, variables **yy1** through **yy4** provide sales revenues at the next level up. For instance **yy1** is the sum of drugs 1-8, i.e., all drugs in the MPH class.

#	Variable	Label
26	rr	Total Revenue (All Drugs)
27	y1	y1: Total Revenue in MPH-IR
28	y2	y2: Total Revenue in MPH-ER
29	y3	y3: Total Revenue in MPH-OROS
30	y4	y4: Total Revenue in MAS-IR
31	y5	y5: Total Revenue in MAS-ER
32	y6	y6: Total Revenue in DEX-IR
33	y7	y7: Total Revenue in DEX-ER
34	y8	y8: Total Revenue in OTH
35	yy1	yy1: Total Revenue in MPH
36	yy2	yy2: Total Revenue in MAS
37	yy3	yy3: Total Revenue in DEX
38	yy4	yy4: Total Revenue in OTH

TABLE 2. Revenues by group of drugs and total

Note also that yy1 is equal to sum of y1, y2 and y3 and similarly, in other groups. Also, per the tree above, y8 is the same as yy4 and rr is the sum of y1-y8 (or equivalently of yy1-yy4).

Sales data by individual drugs 1-17 (Variables 39-107). The next set of variables provide the sales (i.e., revenues), quantities, prices and log prices of 17 drugs (4×17 variables). The variables are named as follows.

#	Variable	Label
39-55	r1-r17	ri: Sales of DRUG i (real 2000 \$s)
56	qq	qq: Total Quantity (All Drugs)
57-73	q1-q17	qi: Qnty of Drug i
74-90	po1-po17	poi: Price of Drug i
91-107	lpo1-lpo17	lpoi: Log Price of Drug i

Note that the variable label provides the name of the drug rather than just a letter ‘i’.

TABLE 3. Sales data by individual drugs

Relative, Average and Absolute Shares (Variables 108-182). The next set of variables are shares of sales of an individual drug within a group. The relevant groups are as given in the earlier figure: for instance, for drug 1 (Ritalin), we want to know the share of sales within the MPH-IR group, thus $s1 = r1/y1$. Briefly, the variables **s1-s17** are relative shares of a drug within class M-F (molecule-form), variables **u1-u8** are shares of forms within molecules, **m1-m4** are shares of sales by molecules within all drugs. Similar names series with an additional o in the name (**so1-so17**, **uo1-uo8** and **mo1-mo4**) are the average values of the same variables where the average is over all areas (counties given by FIPS codes) by year. Finally, the absolute shares are given by the sequence **sr1-sr17**, where for instance, $sr1 = r1/rr$. Note also that since there is only one product in some of the groups, the relative share is by construction equal to one, specifically, $s8 = s11 = s17 = u8 = 1$. The variables are described in the table below.

#	Variable	Label
Relative Shares		
108-124	s1-s17	si: Relative Share of Drug i (in class M-F)
	s1-s3	si: Relative Share of Drug i (in MPH-IR)
	s4-s7	si: Relative Share of Drug i (in MPH-ER)
	s8	si: Relative Share of Drug i (in MPH-OROS)
	s9-s10	si: Relative Share of Drug i (in MAS-IR)
	s11	si: Relative Share of Drug i (in MAS-ER)
	s12-s14	si: Relative Share of Drug i (in DEX-IR)
	s15-s16	si: Relative Share of Drug i (in DEX-ER)
	s17	si: Relative Share of Drug i (in OTH)
125-132	u1-u8	uf: Relative Share of Form f (in molecule M)
	u1-u3	uf: Relative Share of Form f (in MPH)
	u4-u5	uf: Relative Share of Form f (in MAS)
	u6-u7	uf: Relative Share of Form f (in DEX)
	u8	uf: Relative Share of Form f (in OTH)
133-136	m1-m4	mi: Relative Share of Molecule m (in ALL drugs)
Mean Values of Relative Shares		
137-153	sol-sol17	Average values of s1-s17 over all FIPS by Year
154-161	uo1-uo8	Average values of u1-u8 over all FIPS by Year
162-165	mo1-mo4	Average values of m1-m4 over all FIPS by Year
Absolute Shares		
166-182	sr1-sr17	Absolute Share (of revenue) of Drug i

Note that the variable label provides the name of the drug rather than just a letter 'i'.

TABLE 4. Relative Shares of Sales by Groups

Price Instruments (Variables 183-216). Instruments for the prices were generated using the estimated model parameters and are correlated to marginal costs of the individual drugs.

#	Variable	Label
183-199	poz1-poz17	pozi: Instrument for price of Drug i
200-216	lpoz1-lpoz17	lpozi: Instrument for log(price) of Drug i

Note that the variable label provides the name of the drug rather than just a letter 'i'.

TABLE 5. Price Instruments

Price Indexes (Variables 217-242 & 243-268). Thirteen price indexes, one for each segment, starting from the lowest all the way to the top, were computed using Stone price index formula – share weighted average of log price, i.e.,

$$\ln P = \sum_j s_j \ln(p_j)$$

for some relevant group of drugs j in a given class. The relevant groups are as given in figure 1 (or see Table 2) and represent price indexes from 8 lower level segments (one for each segment representing forms within a molecule), 4 middle level price indexes for the molecules (using share

of forms u_j within the molecule as the weighting variable and the lower level 8 price indexes), and one upper level price index for all drugs (using the share of molecules m_j in the formula above as the weight and taking weighted average of the four molecule price indexes).

Further, the same indexes were also computed using the fixed shares (average over all counties over by years), i.e., by using as weights so_j , uo_j and mo_j in the Stone price index formula above. Thus a total of 26 price indexes are provided in the data set and are listed in the table below.

#	Variable	Label
Price Indexes for Forms within Molecules (fixed shares)		
217	lpoi1	lpoi1: wt. avg. lpo1-lpo3 – wt fixed shares (PI for MPH-IR)
218	lpoi2	lpoi2: wt. avg. lpo4-lpo7 – wt fixed shares (PI for MPH-ER)
219	lpoi3	lpoi3: wt. avg. lpo8-lpo8 – wt fixed shares (PI for MPH-OROS)
220	lpoi4	lpoi4: wt. avg. lpo9-lpo10 – wt fixed shares (PI for MAS-IR)
221	lpoi5	lpoi5: wt. avg. lpo11-lpo11 – wt fixed shares (PI for MAS-ER)
222	lpoi6	lpoi6: wt. avg. lpo12-lpo14 – wt fixed shares (PI for DEX-IR)
223	lpoi7	lpoi7: wt. avg. lpo15-lpo16 – wt fixed shares (PI for DEX-ER)
224	lpoi8	lpoi8: wt. avg. lpo17(a,b,c,d) – wt fixed shares (PI for OTH)
Price Indexes for Molecules (fixed shares)		
225	lpomi1	lpomi1: wt. avg. lpoi1-lpoi3 – wt fixed shares (PI for MPH)
226	lpomi2	lpomi2: wt. avg. lpoi4-lpoi7 – wt fixed shares (PI for MAS)
227	lpomi3	lpomi3: wt. avg. lpoi6-lpoi7 – wt fixed shares (PI for DEX)
228	lpomi4	lpomi4: wt. avg. lpoi8-lpoi8 – wt fixed shares (PI for OTH)
Price Index for all drugs (fixed shares)		
229	lpoi	lpoi: wt. avg. lpomi1-lpoi4 – wt fixed shares (PI for ALL)
Price Indexes for Forms within Molecules (actual shares)		
230	lpoi1	lpoi1: wt. avg. lpo1-lpo3 – wt actual shares (PI for MPH-IR)
231	lpoi2	lpoi2: wt. avg. lpo4-lpo7 – wt actual shares (PI for MPH-ER)
232	lpoi3	lpoi3: wt. avg. lpo8-lpo8 – wt actual shares (PI for MPH-OROS)
233	lpoi4	lpoi4: wt. avg. lpo9-lpo10 – wt actual shares (PI for MAS-IR)
234	lpoi5	lpoi5: wt. avg. lpo11-lpo11 – wt actual shares (PI for MAS-ER)
235	lpoi6	lpoi6: wt. avg. lpo12-lpo14 – wt actual shares (PI for DEX-IR)
236	lpoi7	lpoi7: wt. avg. lpo15-lpo16 – wt actual shares (PI for DEX-ER)
237	lpoi8	lpoi8: wt. avg. lpo17(a,b,c,d) – wt actual shares (PI for OTH)
Price Indexes for Molecules (actual shares)		
238	lpomi1	lpomi1: wt. avg. lpoi1-lpoi3 – wt actual shares (PI for MPH)
239	lpomi2	lpomi2: wt. avg. lpoi4-lpoi5 – wt actual shares (PI for MAS)
240	lpomi3	lpomi3: wt. avg. lpoi6-lpoi7 – wt actual shares (PI for DEX)
241	lpomi4	lpomi4: wt. avg. lpoi8-lpoi8 – wt actual shares (PI for OTH)
Price Index for all drugs (actual shares)		
242	lpoi	lpoi: wt. avg. lpomi1-lpoi4 – wt actual shares (PI for ALL)

TABLE 6. Price Indexes

Finally, 26 similar instruments were computed for the 26 price indexes using instruments for log prices instead of log prices. These are described in the table below.

#	Variable	Label
Instruments for PI for Forms within Molecules (fixed shares)		
243	lpzi1	lpzi1: wt. avg. lpz1-lpz3 – wt fixed shares (PIz for MPH-IR)
244	lpzi2	lpzi2: wt. avg. lpz4-lpz7 – wt fixed shares (PIz for MPH-ER)
245	lpzi3	lpzi3: wt. avg. lpz8-lpz8 – wt fixed shares (PIz for MPH-OROS)
246	lpzi4	lpzi4: wt. avg. lpz9-lpz10 – wt fixed shares (PIz for MAS-IR)
247	lpzi5	lpzi5: wt. avg. lpz11-lpz11 – wt fixed shares (PIz for MAS-ER)
248	lpzi6	lpzi6: wt. avg. lpz12-lpz14 – wt fixed shares (PIz for DEX-IR)
249	lpzi7	lpzi7: wt. avg. lpz15-lpz16 – wt fixed shares (PIz for DEX-ER)
250	lpzi8	lpzi8: wt. avg. lpz17(a,b,c,d) – wt fixed shares (PIz for OTH)
Instruments for PI for Molecules (fixed shares)		
251	lpzmi1	lpzmi1: wt. avg. lpzi1-lpzi3 – wt fixed shares (PIz for MPH)
252	lpzmi2	lpzmi2: wt. avg. lpzi4-lpzi7 – wt fixed shares (PIz for MAS)
253	lpzmi3	lpzmi3: wt. avg. lpzi6-lpzi7 – wt fixed shares (PIz for DEX)
254	lpzmi4	lpzmi4: wt. avg. lpzi8-lpzi8 – wt fixed shares (PIz for OTH)
Instrument for PI for all drugs (fixed shares)		
255	lpzi	lpzi: wt. avg. lpzmi1-lpzmi4 – wt fixed shares (PIz for ALL)
Instrument for PI for Forms within Molecules (actual shares)		
256	lpzii1	lpzii1: wt. avg. lpz1-lpz3 – wt actual shares (PIz for MPH-IR)
257	lpzii2	lpzii2: wt. avg. lpz4-lpz7 – wt actual shares (PIz for MPH-ER)
258	lpzii3	lpzii3: wt. avg. lpz8-lpz8 – wt actual shares (PIz for MPH-OROS)
259	lpzii4	lpzii4: wt. avg. lpz9-lpz10 – wt actual shares (PIz for MAS-IR)
260	lpzii5	lpzii5: wt. avg. lpz11-lpz11 – wt actual shares (PIz for MAS-ER)
261	lpzii6	lpzii6: wt. avg. lpz12-lpz14 – wt actual shares (PIz for DEX-IR)
262	lpzii7	lpzii7: wt. avg. lpz15-lpz16 – wt actual shares (PIz for DEX-ER)
263	lpzii8	lpzii8: wt. avg. lpz17(a,b,c,d) – wt actual shares (PIz for OTH)
Instruments for PI for Molecules (actual shares)		
264	lpzmii1	lpzmii1: wt. avg. lpzii1-lpzii3 – wt actual shares (PIz for MPH)
265	lpzmii2	lpzmii2: wt. avg. lpzii4-lpzii5 – wt actual shares (PIz for MAS)
266	lpzmii3	lpzmii3: wt. avg. lpzii6-lpzii7 – wt actual shares (PIz for DEX)
267	lpzmii4	lpzmii4: wt. avg. lpzii8-lpzii8 – wt actual shares (PIz for OTH)
Instrument for PI for all drugs (actual shares)		
268	lpzii	lpzii: wt. avg. lpzmii1-lpzii4 – wt actual shares (PIz for ALL)

TABLE 7. Instruments for Price Indexes

State Dummies (Variables 269-317). The data set also provides a series of dummy variables as d1-d49 for each state and Washington D.C. and does not include any information from Alaska and Hawaii.

REFERENCES

Bokhari, Farasat A.S. and Gary M. Fournier, “Entry in the ADHD drugs market: Welfare impact of generics and me-toos,” *Journal of Industrial Economics*, June 2013, 61 (2), 340–393.