DETAILS OF THE DATA ACQUISITION AND PROCESSING FOR THE DATA USED IN:

Chaudhuri, Shubham, Pinelopi Goldberg, and Panle Jia, "Estimating the Effects of Global Patent Protection for Pharmaceuticals: A Case Study of Quinolones in India"

1. BACKGROUND

The data were purchased at a discounted rate from ORG-MARG, an Indian market-research firm. Under the terms of the agreement (under which the discount was negotiated), the data are to be used solely for this research and are not to be circulated.

Pharmaceutical products are distinguished in terms of the:

- the **active pharmaceutical ingredient (API)** or combination of APIs (therapeutically active molecules) contained in the product
- the **company** under whose name the product is sold
- the particular **presentation** of the product, i.e., the combination of:
  - dosage **form** (e.g., tablet, capsule, syrup, injectible, etc.)
  - dosage **strength** (e.g., 100 mg. tablet)
  - **packet size** (e.g., 100 tablet bottle)

Pharmacists maintain inventory records in terms of **stock-keeping units (SKUs)**. A SKU is a unique combination of an API, the particular presentation in which the product is sold, and the company under whose name the product is sold. Thus, for instance, a 100 capsule bottle of 100 mg. capsules of aspirin sold by Bayer would be considered a different SKU from a 250 capsule bottle of 100 mg. capsules of aspirin marketed by Bayer.

The data provided by ORG-MARG were data at the level of individual stock-keeping units (SKUs). For each SKU available in the Indian market, the data provide information on:

- the therapeutic category of the SKU (to the four-digit level)
- the monthly revenues from retail sales and the number of units sold, for each month from January 1999 to December 2000, for each of the four geographic zones of India
- details about the SKU such as dosage form, dosage strength, packet size, launch date, the company and the brand name under which the SKU was sold

2. DATA DESCRIPTION FOR ORIGINAL DATA

No.	Variable	Datatype	Format	Description, example and codes
	Pcode	str12	%12s	Unique SKU ID number Example: “JAL1A0800213
	Time	str6	%6s	Year and month Example: “199901”
	Area	str3	%3s	Geographic zone Example: “002” Codelist: 002 North 003 East 004 West 005 South
	tc2d	str30	%30s	Therapeutic code and description: 2-digit Example: (codes in Annex 1) "J 1 ANTIBIO./ANTIBAC. SYST. "
	tc3d	str30	%30s	Therapeutic code and description: 3-digit Example: (codes in Annex 1) “*J 1L QUINOLONES”
	tc4d	str30	%30s	Therapeutic code and description: 4-digit Example: (codes in Annex 1) “J 1L1 CIPROFLOXACIN ORAL SOLIDS”
	Comd	str25	%25s	Company under whose name SKU is sold Example: “ALKEM LABORATORIES”
	Brdd	str25	%25s	Brand name under which SKU is sold Example: “ALCIPRO”
	Pakd	str25	%25s	SKU packet description Example: “TAB 750MG 10”
	Insym	str6	%6s	Year and month SKU was introduced in Indian market Example: “199405”
	Unit	Double	%10.0g	No. of units of the SKU sold
	Value	Double	%10.0g	Revenue (Rs.) from retail sales of SKU
	Qty	Double	%10.0g	No. of units of the basic dosage form (e.g. tablets) contained in the SKUs sold
	Inqtyy	Double	%10.0g	Total quantity (`00mg.) of API contained in the SKUs sold

3. STEPS IN THE PROCESSING OF ORG-MARG DATA

STEP 1 Data work at the SKU level

a. The price of 1000 milligrams of API for each SKU in Rs. is obtained by dividing 'value' by 'inqtyy', and multiplying 10.

b. The information on API is manually obtained from the following sources:
   i) The variable 'tc4d' from the original data set;
   ii) Various Indian medical/health websites
   iii) The company websites

c. Information on the domestic/foreign origin of the companies is collected from the following sources:
   i) A firm level proprietary data set from 'Center For Monitoring Indian Economy PVT. LTD' (CMIE);
   ii) For firms not in CMIE, various Indian pharmaceutical trade journals are consulted;
   iii) The companies' websites for firms not mentioned in CMIE or the trade journals.

d. Further cleaning of the SKU level data is carried out. See ANNEX 2 for a detailed description.

STEP 2: Data work at the product level

a. We aggregate SKUs over presentations and firms to obtain a newly defined product group which is only indexed by molecule and domestic/foreign status. The product price is the revenue share weighted SKU price. The product revenue is the sum of the SKU revenues.

b. Among the anti-biotic drugs (the two digit therapeutic segment code is 'J 1'), we focus on the 22 largest molecules that accounted for 95% of all anti-biotic sales.

c. There are 40 products (a product is a unique combination of molecule and domestic/foreign) among those 22 molecules: 22 domestic products and 18 foreign products. Two foreign products have very few observations: foreign Cefedroxil has 36 observations and foreign Sparfloxacin has 20 observations. They are not included in the analysis.

d. In our final data sets, there are 38 products in 8 segments: segment Tetracycline, segment Chloramphenicol, segment Ampicillin, segment Cephalosporin, segment Trimethoprim, segment Macrolides, segment Quinolones, and other penicillin drugs. There are 7 product groups in the Quinolones segment: Foreign Ciprofloxacin, Foreign Norfloxacin, Foreign Ofloxacin, Domestic Ciprofloxacin, Domestic Norfloxacin, Domestic Ofloxacin, and Domestic Sparfloxacin. Foreign Norfloxacin has 92 observations while the rest products have 96 observations (four regions for 24 months). We set the revenue shares of Foreign Norfloxacin to 0 for the four missing periods.

4.DATA SETS AND VARIABLES USED IN THE ANALYSIS

Steps 1 and 2 generate two data sets: LowerAIDSData.csv and UpperAIDSData.csv. The two data sets are combined into the matlab data set AIDSData.mat, which is used by the Matlab programs for the estimation and the counterfactual analysis.

a. The first data set LowerAIDSData.csv contains the following variables that describe the products in the Quinolones segment:

id: product id;

zone; year; month;

GrpShr = (product revenue at time t in region r)/(total Quinolones revenue at time t in region r)

LnGrpP = log(product price at time t in region r)

Grp_LnPIdx = Stone price index for the Quinolones segment at time t in region r
= Σ {GrpShr * LnGrpP}

LnQuinXP_PerCapita = log(Stone price-index deflated per capita Quinolones expenditure)

North = dummy for the northern region
East = dummy for the eastern region
West = dummy for the western region
South = dummy for the southern region
Summer = dummy for the summer season
Monsoon = dummy for the Monsoon season
Winter = dummy for the winter season

SKUN = the number of SKUs in each product at time t in region r

SKUP1 = price at time t in region r for the largest SKU in the product
SKUP2 = price at time t in region r for the second largest SKU in the product
SKUP3 = price at time t in region r for the third largest SKU in the product
SKUP4 = price at time t in region r for the fourth largest SKU in the product
SKUP5 = price at time t in region r for the fifty largest SKU in the product

LnTotalXP = log(Stone-price-index deflated total antibiotics expenditure at time t in region r)

b. The second data set UpperAIDSData.csv contains the following variables that describe the 8 segments of anti-biotics:

SegDesc: the name of the segment

Tc3d: the three-digit therapeutic code

Zone; year; month;

SegShr: the revenue share of each segment at time t in region r among all antibiotics;

LnSegP = log(segment price at time t in region r), where the segment price is the revenue share weighted product price

Seg_LnPIdx: Stone price index for all anti-biotics at time t in region r
$= \Sigma$ {Segshr * LnSegP}

LnXP_PerCapita = log (Stone-price-index deflated per capita antibiotics expenditure at time t in region r)

5. ESTIMATION AND SIMULATION

There are 24 MATLAB program files used in the estimation and the simulation (3 main programs and a total of 21 subroutines). To run the programs please use the matlab data set AIDSData.mat.

a. The main program is called ‘TRIPS_Main.m’ and it calls 17 subroutines P1_*.m to P17_*.m. The main program first obtains the coefficient estimates for both the upper stage and the lower stage, and then performs the counterfactual analysis for various scenarios. To obtain the standard errors, it generates 300 bootstrap samples and carries out the estimation and simulation for each of these 300 bootstrap samples.

The main program ‘TRIPS_Main.m’ generates 9 tables -- table 6: demand elasticities; table 7: marginal cost and markup; table 8: consumer welfare loss; table 9: price changes in the counterfactuals; table 10: forgone domestic profit; table 11: total welfare loss; table 12: profit gain to foreign producers, table A2: coefficients for the lower stage; and table A3: coefficients for the upper stage.

b. The second main program is called ‘TRIPS_Appen.m’. It estimates various AIDS specifications, including: using both regional and seasonal dummies for the lower stage; estimating the lower stage using OLS; not including the SKU prices in the instrument list when estimating the lower stage.

The program generates 2 tables -- table A4: demand elasticities in different seasons for the northern region, and table A5: demand elasticities using OLS coefficients.

c. The last main program is called ‘TRIPS_Appen2.m’. It is similar to ‘TRIPS_Main.m’, except that it uses the upper bound for marginal cost in the counter-factual analysis. The program produces table A6: estimates for the welfare loss assuming MC=P.

d. Tables 1-5 and table A1 are descriptive statistics of the original SKU level data. For a description of tables 6-12 and tables A2-A6 see a)-c) above.