

SUPPLEMENT TO “LIMITED INFORMATION AND ADVERTISING  
IN THE U.S. PERSONAL COMPUTER INDUSTRY”:

A. DATA DESCRIPTION

(*Econometrica*, Vol. 76, No. 5, September 2008, 1017–1074)

BY MICHELLE SOVINSKY GOEREE

This supplement provides technical details of the data used to estimate the limited information model presented in the main paper. The main paper uses data from four sources, some of which are proprietary or limited use. This supplement discusses the various data sets in more detail and provides variable definitions used in the Fortran estimation program. The publicly available data and Fortran programs are provided in the zip file on this site.

1. DATA SOURCES

THE MAIN PAPER uses data from four main sources: Gartner Inc., Simmons Market Research, Competitive Media Reporting (CMR), and the Consumer Population Survey. The product level data were obtained from Gartner Inc. These data are proprietary and must be obtained on a researcher by researcher basis. I provide more detail below as to which data I used and where I obtained these data. The Simmons data and the CMR Leading National Advertisers data were obtained from different libraries across Virginia and Washington, D.C. The data bases and paper publications that generate these data are available to subscribing libraries, but are not posted publicly. I provide detail below regarding which data bases and publications I used to generate the data. Data on the distribution of consumer characteristics was obtained from the Consumer Population survey. These are publicly available and are described in Section 2 and made available on this site. I now discuss the proprietary and limited use data in turn.

The product level data are proprietary and were provided by Gartner Inc.’s Dataquest. Each quarter, Gartner’s Dataquest surveys personal computer systems vendors to estimate quarterly unit shipments and sales. The data used in the main paper were those used to generate the “Personal Computer U.S. Quarterly Shipments Reports” issued by Gartner every quarter. The data consist of vendor (manufacturer), brand, CPU type, CPU speed, and form factor for all PCs sold. For a current link to data from Gartner’s Dataquest, see

<http://www.gartner.com/it/products/research/dataquest.jsp>.

The advertising data were collected from paper publications of Competitive Media Reporting’s *LNA/Multi-Media Publication*, which includes quarterly ad expenditures across ten media. The ten media include magazines, Sunday magazines, newspapers, network television, spot television, cable television,

syndicated television, network radio, national spot radio, and outdoor. The *LNA/Multi-Media Publication* is available from subscribing libraries. More recent “Ad Expenditure Summary” data are made available online (for a reduced fee) for academic research. For information, see

<http://www.tns-mi.com/prodAcademic.htm>.

The consumer level data come from the *Survey of Media and Markets* conducted by Simmons Market Research Bureau. Simmons collects data on consumers’ media habits, product usage, and demographics from over 20,000 households annually. These data are available through many libraries that subscribe to the Simmons CD-ROM data base. These data are typically accessed through a stand-alone PC in the library. Choices is the software used to search the data base. For information on how to use the Choices software, see

<http://library.georgetown.edu/bic/help/choices/simmain.htm>.

## 2. DATA VARIABLE DEFINITIONS

The main program reads in data from eight files. Notes about the data files and data variable definitions associated with each file follow. Note that not all vendors sell in all quarters.

*CPSMAIN.TXT*: These data consist of a sample of “individuals” from the March Consumer Population Survey for 1996, 1997, and 1998. In each year I drew a sample of 3000 observations for individuals who had an annual income greater than or equal to \$5000. Quarterly income data were constructed from annual data and were deflated using the Consumer Price Index from the Bureau of Labor Statistics (see deflator.txt).

*DEFLATOR.TXT*: These data consist of the Consumer Price Index from Bureau of Labor Statistics and the number of U.S. households obtained from the U.S. Bureau of the Census Current Population Reports.

*VMAIN.TXT*: These are aggregated firm level data. Each firm is assigned a vendor number (in parentheses). The firms are Acer (1), Apple (2), Ast (3), Att (4), Compaq (5), Dell (6), DEC (7), Epson (8), Gateway (9), HP (10), IBM (11), Micron (12), NEC (13), Packard-Bell (15), and Texas Instruments (17).

*SALESMAIN.TXT*: These consist of model-specific data.

*ADMAIN.TXT*: These consist of advertising expenditures. There are some models that had advertising data, but no corresponding sales (e.g., Acerpower). These advertising expenditures are included in the ad matrix. For this reason

there are more models in the ad data than in the sales data. To match ad to sales data use idmodel (or idprod to match for a specific model/quarter combination). Use idprod to match ad expenditure to sales data.

*CORPTXT*: These consist of firm advertising expenditures.

*CONSMAIN.TXT*: These are combined data from three years of samples of 6700 observations per year from the Simmons data. All data were previously deflated.

*SALESAGE.TXT*: These are sorted by time. Use idprod to match to sales-main.txt data.

Data File	Variable	Variable Definition
cpsmain.txt	ns	Number of draws
	nqrt	Number of quarters
	ns_t	Number of consumers in period $t = ns/nqrt$
	nchar	Number of consumer demo characteristics
	year	96, 97, 98
	d_i	Vector of demographic characteristics that enter mu in utility function These are demeaned and include: Dummy for aged 30 to 50, household size, dummy for income below \$60,000, Dummy for income above \$100,000, dummy for white male, Dummy for high school graduate
deflator.txt	d_ad	Vector of demographic characteristics that affect phi These include: Dummy for aged 30 to 50, dummy for aged above 50, dummy for married, household size, dummy for income below \$60,000, dummy for income above \$100,000, dummy for white male, dummy for high school graduate, dummy for 1 to 3 years of college, dummy for college degree, years of education if <12, time dummy
	yyq	Year and quarter starting with 95:1 to 2000:3 then annual (95 to 2000)
vmain.txt	cpi	Deflator
	hhs	Number of households (in thousands)
	nqrt	See above
	nven	Total number of vendor specific sales over all quarters
	ven	An id variable of the format vvyyq, where the first 3 digits are the vendor number (other PC vendor is number 100), yy is year, q is quarter (the time period yyq ranges from 96:1 to 99:4) Vendor number
	vqrt	Number of models of pcs sold by each vendor
	vt dum	Quarter (1 = 96:1, . . . , 12 = 98:4)
	vsal	Time dummy variables for year and quarter
	s vt	Total units sales
	s_vt	Market shares (sales/#households in thousands)

(Continues)

Data File	Variable	Variable Definition
vmain.txt (continued)	vprice	Average price
	vad	Newspaper ad expenditures Magazine ad expenditures Television ad expenditures Other medium expenditures
salesmain.txt	nqrt	Number of quarters
	nprod	Total number of products (models) This corresponds to $nprod\_t$ *appropriate quarter
	nprod_t	Number of models sold in period $t$
	nmods	Number of models available (each model counted only once)
	nk	Number of product characteristics
	nnlk	Number of characteristics in xnl
	idprod	An id variable of the format vvbbbtssfyqq, where the first 11 digits are a unique identifier for each model, vv is the vendor number, bbb is the brand, tt is the CPU-type, ss is the CPU speed, f is the form factor, yy is year, q is quarter
	idmodel	An id variable of the format vvbbbtssf
	id	An id variable of the format nnnn, where nnnn = 1-767 represent models
	ven	Vendor number
	brd	Brand number
	cput	CPU type
	ff	Form factor
	x	Constant CPU speed/800 Dummy variable for CPU type = Pentium 1, Pent 2, Pent Pro Dummy variable for form factor = laptop Vendor dummies Quarter
	xnl	Constant CPU speed Dummy variable for CPU type is Pentium (I, II, or Pro) Dummy variable for form factor = laptop
	qrt	Quarter
	tdum	Time dummy variables for year and quarter (leave out 96:1)
	w	Constant ln CPU speed (not used) Dummy variable for CPU type = Pentium 1, Pent 2, Pent Pro Dummy variable for form factor = laptop Quarter
	sal	Total units sales
	s_jt	market shares for inside goods (sales/number of households in thousands)
admain.txt	price	Average price
	idprod	See above
	idmodel	See above
	nqrt	See above

(Continues)

Data File	Variable	Variable Definition	
admain.txt (continued)	nall	Total number of products (models) with positive advertising over all quarters	
	nall_t	Number of models advertised in period $t$	
	nprod	See above	
	xchar	Ven, brd, cput, cpus, ff, pent 1, pent 2, pentpro, laptop, vendor dummies	
	xad	vendor numbers 1, ..., 15 (in these data 14 = Packard-Bell and 15 = Texas) Total advertising done by firm CPUS peed/800	
	ad	Newspaper ad expenditures Magazine ad expenditures Television ad expenditures Other medium expenditures	
	ad_gr	Total group ad expenditures Total newspaper group ad expenditures Total magazine group ad expenditures Total TV group ad expenditures Total other group ad expenditures	
	corp.txt	nqrt	Number of quarters
		nven	Total number of vendors with firm ad over all quarters
		nven_t	Number of vendors with corporate advertising in period $t$
consmain.txt	corp	Corporate advertising	
	cchar_i	Demeaned consumer characteristics for 96 and 97 These include age2, household size, inclow, inchigh, male*white, eduhsgrad (see previous documentation)	
	inc	Deflated income	
	pcnew	Dummy for bought a PC in last year	
	yr	Year	
	id	Individual id where the first digit is the year of the sample	
	demo	age2, age3, dummy for married, household size, inclow, inchigh, male*white	
	cchar	Demo variables and eduhsgrad, eduad, edubs, edusp, 96 dummy, 97 dummy	
	cchar_ad	Demo variables and eduhsgrad, eduad, edubs, edusp, constant, eduhs (edu $\geq$ eduhsgrad)	
	venown	Vendor if own a PC	
	quintl	Magazine quintile Newspaper quintile TV quintile Other quintile	
	salesage.txt	age	Number of quarters idmodel has been on sale (=1 if first time occur)

*Dept. of Economics, Claremont McKenna College, Claremont, CA 91711, U.S.A. and University of Southern California, Los Angeles, CA 90089, U.S.A.; michelle.goeree@gmail.com.*

*Manuscript received July, 2002; final revision received January, 2008.*