Brief description of the data file "SELEC42.DTA"

By Victor Aguirregabiria
January 2002

1. The Company

The data in this paper comes from a supermarket chain that operates in northern Spain. The name of the supermarket chain is Sebastian de la Fuente S.A. Between 1989 and 1996 this company was owned by the bank Banco Bilbao-Vizcaya (BBV). In June 1996 it was acquired by Eroski, which is one of the leaders in the Spanish food distribution industry. This data set covers the period between January 1990 and May 1992, when the company was part of BBV.

The company headquarters and central warehouse is located in Etxebarri, in the Vizcaya province. During this period the company had almost one hundred outlets, most of them located in Vizcaya. This data set comes from the central warehouse. This warehouse orders new deliveries to manufacturers/wholesalers, stores many of the products, sends to the outlets the orders that they make, and decides upon retail prices. The company also supplies products to independent clients (which do not belong to the chain), most of them being small groceries. During this period, all the company outlets were small supermarkets.

Every Monday, the headquarter sends a list of products and prices to its own outlets and to the independent clients. The outlets place orders to the warehouse by phone or through the terminals of a computer network. Orders are typically delivered within the next twenty-four hours. If the central warehouse runs out of an ordered product, this is communicated to the outlet.

Since January 1990 the company has maintained two databases with information on orders, deliveries, inventory holdings, and prices for each item. In one of these databases the information is recorded in continuous time, but it is kept for only one month. At the end of the month a program calculates monthly aggregates/averages for each item and records these values in a second database that contains monthly information. The data in this paper comes from the monthly database between January 1990 and May 1992 (29 months).

The warehouse holds inventories for about half of the items that the supermarket sells. For the other half, the manufacturers or wholesalers deliver directly the product to the outlets of the chain. For instance, most of the varieties of Coca-Cola and Pepsi-Cola are not stored in the central warehouse of the supermarket chain. The data file "SELEC42.DTA" contains information on the 3745 articles (SKU's) for which the central warehouse maintains inventories.

This data set is NOT a "Consumer Scanner Data". During the sample period, this company did not have scanners in their outlets.

2. General Features of the Data File

- Panel of 3745 articles (SKU's) over 29 months.
- Monthly frequency of observations.
- An observation corresponds to one article in one month
- Firm (Supermarket Chain): Sebastian de la Fuente S.A.
- Period: January-1990 to May-1992 (29 months)
- The warehouse of this supermarket chain maintains inventories for all these 3745 SKUs.

3. Variables

NAME	TYPE	DESCRIPTION
dart	string-28	Description of the article (SKU) in Spanish. It is exactly the same description as in the firm's database. It includes information on: type of product, brand name; and size/weight.
subgrupo	integer	Identification number for the "3-digits" classification of SKU's.
producto	long	Identification number for the "5-digits" classification of SKU's.
cprov	long	Code of the supplier (i.e., manufacturer or wholesaler) that provides the SKU's to this supermarket chain.
numart	integer	Code of the SKU.
month	float	Month: 1 to 29. "1" is January-1990 and "29" is May-1992.

NAME	TYPE	DESCRIPTION
c	float	Unit wholesale price of the SKU. Measured in pesetas. It has been calculated as the ratio between the value of orders and the amount of orders in physical units. When this price doesn't change between two orders, it has been assumed constant over time.
ptp	float	Unit retail price of the SKU under sale promotion. Measured in pesetas. It has been calculated as the ratio between the value of sales under promotion and the amount of sales under promotion in physical units.
p	float	Unit retail price of the SKU. Measured in pesetas. It has been calculated as the ratio between the value of sales and the amount of sales under promotion in physical units.
ofeun	float	Sales, in physical units, under promotion.
dfo	float	Returns, in physical units, from customers (outlets) to the central warehouse.
tp	byte	Dummy for the existence of a sales promotion
s	float	Inventories, in physical units, of the SKU in the central warehouse at the end of the month.
peso	float	Size of the product (Weight or number of units). Same measurement units for SKU's in the same 5-digit producto
sf	byte	Dummy for "own brand", i.e., Sebastian de la Fuente brand.

NAME	TYPE	DESCRIPTION
dtm	float	Returns, in physical units, from the central warehouse to manufacturers.
q	float	Deliveries, in physical units, to the central warehouse from manufacturers.
x	float	Sales of the SKU in physical units. Deliveries from the central warehouse to the outlets.
dem	float	Demand from outlets in physical units. Orders that outlets place to the central warehouse.
exde	float	Excess of demand from outlets: exde = dem – x
markup	float	Markup of the retail price over the whosale price: (p - c)/c
timeacti	byte	Number of months that the SKU has been active in the central warehouse (active means sales>0).
qout	float	Probability of stockout. qout = exde / dem