

Consumer Panel Data: A ZUMA research service

By Georgios Papastefanou

1. Introduction

In modern societies research on private consumption has got a pervading relevance for marketing as well as for political administration and planning. Socio-economic transformations and challenges like sustainable growth or demand for innovations underline the necessity of empirically grounded knowledge on consumer behavior. On the other side consumer behavior patterns seem to be highly fragmented and diversified as social reality is transformed by lifestyle individualization and household structure differentiation. Social sciences have acknowledged this by theoretical perspectives (Friedrichs 1998). But nevertheless most lifestyle studies concentrate on consumer attitudes and value orientations (Hartmann 1999). Only few studies combine behavioral data with data on household composition and as well as with data on attitudes. Actually, no standard micro data base for research on consumption behavior and its psychological and sociological correlates is available.

On the other side, non-academic market research institutes like GfK in Nuremberg, collect detailed information and provide information services on consumer behavior (Twardawa, Wildner 1998). Special panel services run consumer panels producing continuously longitudinal behavioral data on consumption of individuals and households. Aimed at supporting individual companies in marketing their specific products, consumer panels continuously are measuring the product flow into the household over years, what means that they are measuring the household's purchase event histories. These data are combined on a microdata level with socio-economic and demographic information on the household as well as with survey data on attitudes of specific reference persons in the household. So consumer panel data are of high potential for social science consumer analysis, but until now were not accessible generally.

In 1997 ZUMA took the initiative to establish an easy scientific access to consumer panel data. GfK, who runs a long established consumer panel, supported this initiative and gave the data to ZUMA for preparation and distribution to an interdisciplinary working group. After having proofed

the utility and manageability of the prepared consumer panel data in a conference (Papastefanou et al. 2001), GfK agreed in ZUMA's role in preparing and distributing the panel data as the ZUMA Consumer Panel Data of 1995.

This paper describes the ZUMA Consumer Panel Data of 1995: (1) It gives an outline of the contents of the data. (2) It reports on selected topics using the potentiality of data. (3) It highlights the special feature of purchase history data and (4) it informs on the easy access to the micro data for scientific research and education.

2. Information in the data

The ZUMA scientific use file of consumer panel data is based on the GfK ConsumerScan household panel. In this consumer panel GfK is running two independent samples with about 12000 households. They are designed as quota samples, quoted by federal state, administrative district, size of town and household size, with marginal distributions adapted to Mikrozensus distributions. The data, which are delivered by ZUMA as scientific use files are subsamples, containing those households, which in 1995 reported continuously their purchase diaries. These two subsamples comprise 9064 households.

The ZUMA version file consists of two main types of data: a) data on each household's product purchases in 1995 and b) data on the socio-economic and demographic structure of the households at the beginning of 1994, 1995 and 1996, as well as data on attitudes towards nutrition, environment and consumption of the person running the household in October 1994 and October 1995. These two types of data can be merged by a household identification number.

Household purchases in 1995

Core of the consumer panel data collection are a household's purchases all over time with precise information on the timing of every purchase. In past, household budget diaries were used for measurement. Households used a booklet, in which they note all the purchases plus some of their characteristics.¹ The ZUMA consumer panel data of 1995 are based on paper-and-pencil household budget diaries.

¹ Because of technological and product diversity the paper and pencil approach is replaced by home screening method of data collection. The participating households do have a manual screening device, by which they can collect the information of EAN code. The data then automatically are stored in a base station and are transferred later by telephone to the GfK data bank of the consumer panel. The home screening method was installed at GfK in 1998.

In a methodological perspective one can characterize the data as event history data, which are not gathered retrospectively, but simultaneously with the process of product inflow into the household. The data are organized in asymmetrical data sets, in which the purchased product is unit of observation. So, for each household there may be an unequal number of purchase events. Their household assignment is achieved by a household identification number on every purchase record. The GfK ConsumerScan household panel is focused on fast moving consumer goods, especially on packed nutrition products and products of household convenience (see table 1).

For each purchased product following variables are available:

- Date of purchase
- weekday of purchase
- product category (measured in total purchases in 81 product categories, see table 1)
- product subcategory (like flavor, taste etc.)
- type of retailer
- number of purchased products
- type of price (standard vs. special price)
- total quantity purchased
- total amount of expenses
- duration since last purchase
- sort id of purchase (the sort id indicates the sequence number of every purchase in a household)
- specific characteristics of the product, like type of package etc.

Socio-economic and demografic situation of the household

The socio-economic and demografic situation of the household is measured by a paper and pencil survey at the beginning of each year respectively. The following household characteristics are available in the ZUMA dataset:

- federal state of residence
- size of community
- age of the household leading person
- number of children (up to the age of 6, under the age of 14, under the age of 18)

Table 1

Product categories of the ZUMA consumer panel data file

No of product category	Sample 1	Sample 2	Label of product category	No of purchased products in 1995
0	x	x	Window / carpet / toilet-cleaner	25.059
1	x	x	Only sample 1: tomato puree, only 6: pastasaucses, both: Ketchup, spicesaucses	47.651
2	x	x	Only sample 1: mayonese, tartar sauce, both: dressing	27.527
3	x	x	Special detergents	22.304
4	x	x	Detergents for dishes	34.556
5	x	x	Detergents for the household	20.940
6	x	x	Only 1: denture detergents both: toothbrushes / mouthwash	26.818
7		x	Tinned vegetables	63.275
8	x		Milk	204.339
10	x	x	Water softener	29.542
11	x	x	Dental care products	55.497
12	x	x	Pure coffee (roasted)	143.194
13	x	x	Pure coffee-instand	31.576
14		x	Salt	13.100
16		x	Poultry	33.907
17	x	x	Frozen food	230.841
18	x	x	Tea	47.426
19	x	x	Cocoa	17.680
20	x	x	Spirituos beverages	59.129
21	x	x	General detergents	41.225
22		x	Fats	233.124
23	x		Mustard / horseradish	15.055
24		x	Prefabricated cake	23.931
25		x	Soap / washing lotion	13.105
26	x	x	Sparkling wine	30.127
27	x	x	Cider	3.229
28	x	x	Ground care	4.536
29	x	x	Bathing additives	33.206
30	x	x	Finished potato products	39.077
31	x		Pudding / dessert	33.399
32	x		Household tissues	12.265
33	x	x	Beer	131.245
34	x	x	Vermouth / appetizer	2.767
35	x	x	Wine / mulled wine	27.614
36	x	x	Alcoholfree uncarbonated beverages (fruit juice)	187.349
37		x	Winned cabbage	48.502
38		x	Crispbread	17.367
39	x		Shoe- and leather care	3.584
40	x	x	Salt for the dishwasher	4.703
42	x	x	Coffee stuff (like Caro)	3.564
43		x	Delicatess salads	41.375

Table 1 (continue):

No of product category	Sample 1	Sample 2	Label of product category	No of purchased products in 1995
44	x	x	Sherry / Port wine	3.343
45	x		Pudding powder	14.454
46	x	x	Alcoholfree carbonated beverages (no mineral water)	155.447
47	x		Cream cheese / soft cheese	144.046
48	x	x	Detergents for curtain	6.364
50	x		Cream	74.009
51	x	x	Milk for coffee	111.247
52	x		Air fragrances	2.683
53	x	x	Beverages with wine	5.292
54		x	Gingerbread etc.	19.778
55	x		Napkins	4.432
56	x		Lights, matches etc.	2.730
57		x	Vinegar	10.925
59	x		Food care	1.113
64		x	Sweet pastries	59.473
66	x	x	Animal food / cat litter (not prefabricated: only 6)	123.133
68		x	Rusk	7.355
71		x	Dry ready-to-serve meals	9.369
72		x	Products for plants	3.598
73	x		Curd / junkets	94.575
74		x	Salty pastries	19.174
75	x		Filter paper	12.648
76		x	Instant soups	6.062
77		x	Canned meals	11.244
78	x		Yoghurt	119.270
79		x	Flour	25.886
80	x		Cleaning sponge	9.417
81	x		Cheese	96.766
82	x	x	Detergents for the bath	6.278
83		x	Baking powder / yeast	19.084
84	x	x	Mineral water	174.470
85	x		Sandwich paper	18.756
86	x	x	Icecream	59.108
87		x	Ready-to-bake blends	6.875
89		x	Rice	16.158
90	x		Cereals (Conflakes)	33.256
91		x	Pasta	53.201
93		x	Ready-made menus	4.641
94		x	Roasted nuts etc.	58.008
99	x		Toilet paper	3.245

- occupational status of the household leading person
- current occupation of the main income earner
- occupation of the main income earner
- former occupation of the main income earner
- educational status of the main income earner
- size of house
- equipment of the household (washing machine, dishwasher, microwave, VCR)
- garden ownership / -use
- size of garden
- pets (cat, dog, budgie)

Socio-economic and demographic data in the ZUMA dataset are available for three waves of the ConsumerScan panel, namely of 1994, 1995 and 1996. The files of the different years can be merged by a household identification number, allowing for analysis of individual household structure changes.

Attitudes of the person running the household

Surveys on attitudes of specific persons in the household are part of the consumer panel. Under the assumption, that persons, who are running the household represent a kind of gate-keepers, they are the reference persons for the attitudinal survey. The ZUMA file version contains two attitude surveys of Octobre 1994 and Octobre 1995. They can be merged with the product purchase history data by a common household identification number.

The attitude measures cover the following items:

Attitudes towards nutrition: Items on medical health, natural, joy of discovering, german products, convenience orientation, slimness orientation, plain fare, full grain nutrition, superior savour, freshness orientation, pro branded goods, pro vitamins / minerals, uncritical style of nutrition

Attitudes towards aspects of daily life: Items on tendency to innovate, traditional living, experience-orientation, nostalgia, mistrust towards new products, quality-orientation, convenience-orientated cooking

Attitudes towards environment: Items on ecological awareness, environmental behaviour, environment and mobility, state and industry

Price consciousness of the household leading person: This variable contains a composite index which is based on ratings of three item pairs.

3. Selected topics of analysis

The ZUMA consumer panel data of 1995 offers unique possibilities for consumer behavior analysis. Since this type of data was not easily accessible for social and economic research in past there are only few empirical analyses. Wimmer (1995) for example used the data to estimate types of environmental oriented consumers to check for consistency of behavior and attitudes. By cluster analysis of the attitude data enclosed in the 1993 wave Wimmer (1995) specified two groups of households, those with a high environmentalism score and those with low scores. Then he compared these two groups in relation to the bought amount of household cleaning and shower bath products. He could show that there was a consistency between favorability towards environmentalism and actual purchasing of products, which presumably reduce environmental loading. As the data provided detailed information on specific product types, differentiated results were possible. Behavioral consistency was different for different product categories. High consistency was found in relation to amounts of softeners, sanitary cleaners and bubble bath ingredients. Low consistency was given in relation to water closet cleaners, floor cleaners, universal detergents and shower baths. Similar results could be found for packing materials like nonreturnable bottles, glass and cans for carbonic acid and fruit beverages.

A more recent analysis based on the ZUMA consumer panel data access is that of Thiele and Weiss (2001). The authors use the behavioral data to analyze empirically the determinants of variety in food consumption. Obviously no German analysis was done until now on this issue, the published research was based on US data expenditure survey data from 1972/73, 1977/78 and 1981, the last ones being restricted to urban households in the north-eastern region. Thiele and Weiss (2001) estimated several indexes for the household variety in food consumption for 149 food groups available in the purchase data and related them to several socio-economic household characteristics. They found that variety increases with household income, with the number of children aged between 6 and 18 years. Food diversity is also significantly higher if the family lives in larger cities in East-Germany and if the house-keeping person is not working full-time. Food product variety in single male households is significantly lower than that of female singles.

A ZUMA symposium on the manageability and potential of the consumer panel data revealed a multitude of further relevant research issues. The results are reported in Papastefanou et al. (2001). Following, the topics of some of the contributions will be described shortly.

In one paper, considering meanings of personal and social communication and exclusion by behavioral life style patterns, Lüdtke and Schneider (2001) examine the data for life style strategies that can also be observed in

purchasing everyday consumer products. By including indicators of social position and attitudinal expression they further differentiate behavioral patterns along subjective and social meanings. One of the central findings is that social position and mentality correlates with fast moving consumer good purchases like food, beverages and hygiene products.

Another analysis (Thøgersen 2001) deals with the spread of sustainable behavior patterns of individuals. Thøgersen asks if there is a mechanism of spillover, where a proenvironmental purchasing disposition (expressed in the purchasing of specific goods) is transferred to other environmentally relevant products. Using product information on packaging of purchased products he analyses also if spillover is supported by specific attitudes and values (Thøgersen 2001).

Ohr (2001) relates to the discussion on health risks in modern food production. He asks if purchasing of food products being related to health issues are influenced by attitudes towards nutrition and health. He finds consistency between purchasing and attitudes towards healthy nutrition, also if the involvement is on a medium level. Differentiating the expenses for healthy food products according to age, education and region he underlines that social position, mediated by socialization and cognitive competence, determine everyday life.

Consumer panel data can be used also for critical evaluation of theoretical economic models on price formation and examination of dynamics and variation of retail prices. Fengler and Winter (2001) analyzed the purchases of coffee, as it is a product being bought frequently and having changing prices. Their results show the importance of psychological price thresholds for price settings and changes along changes in costs.

The event history information potential of consumer panel data is utilized in Fachinger (2001), who is interested in risk evaluation by consumers. He argues that individual risk evaluation should be best detected when considering continuity of purchasing behavior. The available information on the timing of a purchase act allows to apply hazard rate models on purchases of coffee, detergents and yeast.

Information on changes in net household income between two waves is used in the analysis of Papastefanou (2001). He examines the hypothesis that consumption of alcoholic beverages (with more than 20 Vol. alcohol) is correlated with household income position. The panel data made it possible to separate income change from income level effects on the bought amount of liquors. By this, empirical support could be found for a positive income level effect, but a negative income change effect.

4. Special potential for behavioral analysis: purchase histories

As mentioned above, the ZUMA consumer panel data set consists of socioeconomic and attitudinal data plus event histories of product purchases. One major potential of this merged data set is the combination of these types of data. In the following the special potential of event history data is highlighted.

Event history data provide the data on changes that may occur at any point of time. There are time related information on the purchase plus its specific attributes (see above), so the dependent variable can be conceived as a variable, which can change its state at any time. Thereby the transition rate framework is applicable, which offers a time-point-related representation for causal effects (Blossfeld/Rohwer 1995). Based on cross-sectional data, only correlational analysis is possible. In correlational analysis causal interpretation is achieved indirectly holding constant correlated variables. But the time ordering of independent and dependent variables, which is available in longitudinal data, is a necessary prerequisite for causal analysis. The event history data on purchases not only fulfill this methodological necessity, but they make possible to estimate adequately the process of product flow into the household. Purchases of products are not isolated acts, but they are part of a dynamic process, in which realized states and changes influence following states and changes. Event history analysis with the transition rate as the core statistical concept offers sophisticated tools for modeling this process. The high analytical power of the purchase data is even fostered by the fact, that these data are not – as usually – gathered retrospectively, but are quasi process produced data. Recall bias, survivor selection bias and misrepresentation bias are less effective.

Of course, analyzing purchase event history data needs some special effort, because the data are organized asymmetrical. But standard statistical analysis programmes like SPSS, SAS, Statistica etc. offer easy applicable routines for data preparation (aggregation of the individual purchase histories, defining competing risks and types of individual changes of purchase strategies/patterns) and event history analysis. More sophisticated special software is also available, like TDA or LIMDEP. ZUMA's service include counseling in this issues.

5. Data access

The ZUMA consumer panel data 1995 is freely available for scientific research and education. Researchers and teachers can get the micro data after sending in a signed terms of agreement and paying a fee of 50 Euro. Detailed

information on the data, on the survey documents as well as the terms of agreement are available available on the CD-ROM and in the web at http://www.gesis.org/Dauerbeobachtung/Einkommen/Daten/Gfk/gfk_index.htm.

6. A look ahead

ZUMA started this service with consumer panel data, because there seemed to be a barrier for economic and social analysis of consumption. One main barrier is due to the lack of data, which are behavioral and dynamic, and which can be connected or fused with data on household structures and on attitudes and values.

The ZUMA consumer panel data 1995 version of the GfK ConsumerScan files is the first dataset of this kind being freely accessible by the scientific community. It is restricted yet, as it is focused on data on fast moving consumer goods. Also the data are from the 1995 wave. The attitudinal data are related to environmental, nutritional and general consumer issues, the household structure information is not that detailed like e.g. the ZUMA demographic variables set.

But this restrictions are not real constraints, as this first time public private cooperation in scientific research is open for further development. It is possible to get more actual data to monitor latest political and economic events, to get also older datasets for analyzing trends, and to insert attitudinal surveys (as long as they do not disturb the panel participation process) on specific issues. Further, the tremendous data production of the GfK consumer panel services offers also data on fresh products, alcoholic beverages, confectionary, personal care, drugs, and communication and services (magazines, cinema, video etc.). There is no principal barrier against the scientific access to these data. If the scientific community proves to be interested and to profit from consumer panel data, the bases for extending the scientific service with consumer panel data is given.

References

- Berger, Roger* (2001), Consumer panel data and rational choice based theories of myopic formation. An empirical analysis, in: Papastefanou, Georgios et al., *Social and Economic Research with Consumer Panel Data. ZUMA Nachrichten Spezial Band 7*.
- Blossfeld, Hans-Peter / Rohwer, Götz* (1995), *Techniques of Event History Modeling. New Approaches to Causal Analysis*, Mahwah, N.J., Lawrence Earlbaum Assoc.

- Boztuğ, Yasmin / Hildebrand, Lutz* (2001), Nonparametric modeling of buying behavior in fast moving consumer goods markets, in: Papastefanou, Georgios et al., Social and Economic Research with Consumer Panel Data. ZUMA Nachrichten Spezial Band 7.
- Fachinger, Uwe* (2001), Continuity or Discontinuity in the decision of the application of income of private households?, in : Papastefanou, Georgios et al., Social and Economic Research with Consumer Panel Data. ZUMA Nachrichten Spezial Band 7.
- Fengler, Matthias / Winter, Joachim* (2001), Price-setting and price-adjustment behavior for fast moving consumer goods, in: Papastefanou, Georgios et al., Social and Economic Research with Consumer Panel Data. ZUMA Nachrichten Spezial Band 7.
- Friedrichs, Jürgen* (Ed.) (1998), Die Individualisierungs-These. Opladen: Leske + Budrich.
- Hartmann, Peter* (1999), Lebensstilforschung. Darstellung, Kritik und Weiterentwicklung. Opladen: Leske + Budrich.
- Lüdtke, Hartmut / Schneider, Jörg* (2001), Can patterns of everyday consumption indicate lifestyles? A secondary analysis of expenditures for fast moving consumer goods and their social contexts, in: Papastefanou, Georgios et al., Social and Economic Research with Consumer Panel Data. ZUMA Nachrichten Spezial Band 7.
- Ohr, Dieter* (2001), Purchasing healthy food in Germany. An empirical analysis of its attitudinal and socio-economic antecedents, in: Papastefanou, Georgios et al., Social and Economic Research with Consumer Panel Data. ZUMA Nachrichten Spezial Band 7.
- Papastefanou, Georgios* (2001), Household's income situation and consumption of alcoholic beverages, in: Papastefanou, Georgios et al., Social and Economic Research with Consumer Panel Data. ZUMA Nachrichten Spezial Band 7.
- Papastefanou, Georgios / Schmidt, Peter / Börsch-Supan, Axel / Lüdtke, Hartmut / Oltersdorf, Ulrich* (eds.) (2001), Social and Economic Research with Consumer Panel Data. ZUMA Nachrichten Spezial Band 7.
- Prester, Hans-Georg* (2001), Consumer Panel Research of GfK, in: Papastefanou, Georgios et al., Social and Economic Research with Consumer Panel Data. ZUMA Nachrichten Spezial Band 7.
- Thögersen, John* (2001), Is a sustainable consumption pattern gradually emerging in Germany? Gradual changes in the sustainability of the consumption pattern analysed by means of a consumer panel, in: Papastefanou, Georgios et al., Social and Economic Research with Consumer Panel Data. ZUMA Nachrichten Spezial Band 7.
- Thiele, Silke / Weiss, Christop R.*, 2001, Variatio Delectat. Consumer Demand for Food Diversity, Working Paper EWP 0102, University of Kiel
- Twarda, Wolfgang / Wildner, Reinhard* (1998), Innovationsforschung mit Paneldaten, planung & analyse, Heft 3, 10 – 15.
- Wimmer, Frank* (1995), Der Einsatz von Paneldaten zur Analyse des umweltorientierten Kaufverhaltens von Konsumenten, in: UWF – Umweltwirtschaftsforum, März, 28 – 34.