

## The ZUMA data file version of the GfK ConsumerScan household panel

Papastefanou, Georgios

Veröffentlichungsversion / Published Version

Sammelwerksbeitrag / collection article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

### Empfohlene Zitierung / Suggested Citation:

Papastefanou, G. (2001). The ZUMA data file version of the GfK ConsumerScan household panel. In G. Papastefanou, P. Schmidt, A. Börsch-Supan, H. Lüdtke, & U. Oltersdorf (Eds.), *Social and economic research with consumer panel data : proceedings of the first ZUMA Symposium on Consumer Panel Data, 5 and 6 October 1999* (pp. 206-215). Mannheim: Zentrum für Umfragen, Methoden und Analysen -ZUMA-. <https://nbn-resolving.org/urn:nbn:de:0168-ssoar-49470-1>

### Nutzungsbedingungen:

Dieser Text wird unter einer Deposit-Lizenz (Keine Weiterverbreitung - keine Bearbeitung) zur Verfügung gestellt. Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### Terms of use:

This document is made available under Deposit Licence (No Redistribution - no modifications). We grant a non-exclusive, non-transferable, individual and limited right to using this document. This document is solely intended for your personal, non-commercial use. All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

# THE ZUMA DATA FILE VERSION OF THE GfK CONSUMERSCAN HOUSEHOLD PANEL

*GEORGIOS PAPASTEFANOU*

## 1. Data structure

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 household. 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 household identification number.

### *Household purchases in 1995*

Core of consumer panel data collection is the history of a household's purchases all over the time with precise information on the purchases's point of time. Usually, household budget diaries were used for measurement. Households voluntarily use a booklet, in which they note all the purchases plus some characteristics.<sup>1</sup> The ZUMA scientific use file consumer data are bases on this type of paper-and-pencil household budet 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

---

Methodologically, the collected data are event history data, which are not gathered retrospectively, but in a kind of measurement by process, in continuous time. By this, the data are organized as an asymmetrical data set, 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 an added household identification number, being constant for each purchase records, as long it does belong to the same household.

The GfK ConsumerScan household panel is focused on fast moving consumer goods, especially on packed nutrition products and products of household convenience. For each purchased product there are available several variables:

- *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*
- *waiting time since last purchase*
- *sort id of purchase (the sort id indicates the sequence number of purchase since the beginning of 1995)*
- *specific characteristics of the product, like type of package etc.*

---

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.

Table 1: Product categories of the GfK ConsumerScan Household Panel as being available in the ZUMA file version

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: pastasauces, both: Ketchup, spicesauces	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	Spirituous 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

Table 1 (continued): Product categories of the GfK ConsumerScan Panels as being available in the ZUMA file version

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
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 fragrant	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

*Table 1 (continued): Product categories of the GfK ConsumerScan Panels as being available in the ZUMA file version*

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

***Socio-economic and demographic situation of the household and attitudes of the person running the household***

The socio-economic and demographic 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)
- 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)

The 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, which makes possible analyzing individual household structure changes.

### *Attitudes of the person running the household*

Surveys on attitudes of specific persons in the household are inserted into the consumer panel.

Assuming, that persons, who are running the household represent some function of gatekeepers, they are the reference persons for the attitudinal survey. The ZUMA file version contains two attitude surveys of Octobre 1994 and Octobre 1995. These data are part of the variable list in the household structure file and can – like them – be merged with the product purchase history data.

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, un-critical 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.

## **2. How to combine purchase history data and household structural data**

As we made clear above, socioeconomic and attitudinal data are organized in one type of file with a rectangular data matrix. This means, that for each observation unit there is on

data record. This type of data we will call household structure data (including the attitudinal data of a reference person). The purchase history data are asymmetrical data files.

The purchase history data of 1995 are organized as separate files for each product category. Be aware, that these files do contain all the purchase of households of both samples, and that some of the socioeconomic demographic or attitudinal data might not be available for both samples.

To combine purchase history data with household structure data, one has to transform the asymmetrical purchase history file into a rectangular data matrix. This can be done by a preselection or by household specific aggregation of purchase variables.

In case of preselection a specific purchase in the order of all purchase will be selected by using the sort id. Thereby it is possible, to condense the asymmetric purchase history into a rectangular household data matrix with the first or the second or the third etc. purchase. It is not possible to choose by this preselection the last purchase done in 1995, because there is not known the sort id of the last purchase in each household.

For household specific aggregation of purchase variables in SPSS AGGREGATE procedure there are different functions. The main logic is to indicate the household context by a break variable (which in our case is represented by the household identification number), which is on each purchase record, but which changes if the household assignment changes. According this household context the purchase variables can be aggregated into one value by using different functions mean, standard deviation, minimum value, maximum value, sum of values, number of purchases for each household (in SPSS terminology these are the number of cases for each value of the break variable). Besides these central tendencies of the household specific distribution of the purchase variables one can also get statistics for parts of the household specific distributions. To get this one has to define a threshold value or an interval and then to compute the percentage or the fraction of purchases being below resp. above the threshold or inside resp. outside the interval.

At last it might be interesting, that by the AGGREGATE procedure one can also run a selection of the first or the last purchase record and its variables.

By aggregating the household's purchase history into one central tendency value one gets a new file containing the household id number plus the aggregated values of the purchase variables. This is a rectangular data matrix which can be merged with the data matrix on household structure and attitudes measures. So, covariances of central behavioral tendencies of the household with its socio-economic structure and the attitudes of its reference person can be analyzed.



## **Einverständniserklärung**

Der Datenempfänger verpflichtet sich durch Unterschrift, nach Erhalt der ZUMA-Daten des GfK-Verbraucherpanels von 1995 zur Einhaltung folgender Bedingungen:

- 1 Der Datenempfänger und alle Personen, denen die Daten zugänglich gemacht werden, verpflichten sich, keine De-Anonymisierungsversuche zu unternehmen und die Daten nicht an andere Personen - außer den Mitarbeitern des Datenempfängers - weiterzugeben oder sie ihnen zugänglich zu machen. Dies gilt auch für modifizierte Daten.
- 2 Die Daten dürfen nur in der eigenen wissenschaftlichen Forschung und der universitären Lehre des Datenempfängers eingesetzt werden. Eine Nutzung für gewerbliche Zwecke ist nicht gestattet.
- 3 Für die Überlassung des Nutzungsrechts an den Daten wird ein Preis von Euro 20.- inkl. MWST\* vereinbart. Der Betrag wird mit meinem Einverständnis von meinem Konto abgebucht.
- 4 In den Forschungsarbeiten, in denen Ergebnisse der ZUMA-Daten des GfK-Verbraucherpanels von 1995 eingehen, wird folgender Quellenverweis aufgenommen:

*Die in diesem Beitrag benutzten Daten entstammen dem Datensatz, den die GfK, Nürnberg als Unterstichprobe des ConsumerScan Haushaltspanels von 1995 ZUMA zur Verfügung gestellt hat. Dieser ZUMA-Datensatz enthält alle Haushalte, für die im Jahr 1995 durchgehend Kaufdaten gesammelt worden sind. Für eine nähere Beschreibung der Verbraucherpaneldaten siehe: Papastefanou, Georgios: The ZUMA scientific use file of the GfK ConsumerScan Household Panel, in: Papastefanou, Georgios / Schmidt, Peter / Börsch-Supan, Axel / Lüdtke, Hartmut / Oltersdorf, Ulrich (eds.), 2001, Social and Economic Analyses of Consumer Panel Data, Zentrum für Umfragen, Meinungen und Analysen (ZUMA), Mannheim.*

---

\* Dieser Preis ist ein Subskriptionspreis. Er gilt in Verbindung mit Erwerb des ZUMA –Spezial Nr. 7 und ist gültig bis 31.12.2001.

Mit meiner Unterschrift verpflichte ich mich die umseitig aufgeführten Bedingungen einzuhalten.

Adresse des Datenempfängers

Ort, Datum

Unterschrift

*Einzugsermächtigung*

Ich erteile hiermit dem Zentrum für Umfragen, Methoden und Analysen (ZUMA) die Berechtigung, den Betrag von Euro 20.- von meinem Konto Nr. .... bei Bankinstitut ..... abzubuchen.

Ort, Datum

Unterschrift

An  
ZUMA -Abt. Einkommen und Verbrauch  
Postfach 122155  
68072 Mannheim

---

## Terms of Agreement

By signing this document the recipient of data complies to the following rules concerning usage of the ZUMA scientific use file of the 1995 GfK ConsumerScan Household Panel Data (in short: ZUMA/GfK Data):

1. The recipient of the ZUMA/GfK Data, as well as all other persons to whom these data are made accessible, are committed not to undertake any attempts to de-anonymize these data. The data must not be made accessible to anyone with the exception of collaborators of the data recipient. This applies also to modified versions of these data.
2. The ZUMA/GfK Data must not be used for purposes other than the data recipient's personal scientific research and academic teaching. It is not allowed to use the data for commercial purposes.
3. The data recipient will pay a fee of Euro 20.-\* for obtaining the ZUMA/GfK Data and being granted the right to use them for scientific purposes as indicated above.
4. All publications reporting findings derived from the ZUMA/GfK Data must refer to the data source by including the following annotation:

*The data used for this analysis are part of a subsample of the 1995 GfK ConsumerScan Household panel data and were made accessible by ZUMA. The ZUMA data set includes all households having continuously reported product purchases during the entire year 1995. For a description of this data set cf. Papastefanou, Georgios, 2001, The ZUMA scientific use file of the GfK ConsumerScan Household Panel, in: Papastefanou, Georgios / Schmidt, Peter / Börsch-Supan, Axel / Lüdtke, Hartmut / Oltersdorf, Ulrich (eds.), 2001, Social and Economic Analyses of Consumer Panel Data, Zentrum für Umfragen, Meinungen und Analysen (ZUMA), Mannheim.*

Name and address of data recipient

Date and signature of data recipient

---

\* This subscription price is effective until December 31, 2001, and only in combination with the purchase of this ZUMA Special Volume No. 7.