Saving and Old Age Provision in Germany (SAVE): Design and Enhancements

By Michela Coppola and Bettina Lamla

1. Introduction

Understanding how households form their long-term saving and investment decisions to shoulder risks not covered by social security systems has been of primary importance in all the countries which, like Germany, introduced major reforms to face the challenges of an aging population. These reforms have shifted longevity and income risks from the state to the individual and therefore require individuals to make their own provisions. About a decade ago, the data situation in Germany for analyzing households' saving and financial behavior was very limited. None of the then existing databases recorded detailed information on both "hard" financial variables (e.g. income, assets, debts) and "soft" sociological and psychological characteristics. Moreover, none of the studies offering detailed financial variables (such as the official Income and Expenditure Survey – EVS) had a longitudinal structure or were conducted within reasonably short time intervals.

To bridge this gap, the Mannheim Research Institute (since 2011 Munich Center) for the Economics of Aging (MEA) initiated the study "*Saving and Old-Age Provision*" (SAVE) in 2001. The main goal of the study was to create a sound empirical base to better understand households' saving behavior and asset choices. It was thus designed to collect at the household level detailed information on income, financial and real assets as well as debt in combination with a rich set of psychological questions, questions on health, expectations and attitudes. Furthermore, it was set up as a longitudinal study to allow monitoring developments over time and studying reactions to macroeconomic and institutional changes.[1]

Although much has been learnt in the last years on the behavioral aspects of handling risk, uncertainty, and long-term planning, the specific challenges posed by the recent socio-economic and political developments such as the

---

[1] For a detailed description of the scientific background of the study and of its design see Börsch-Supan et al. (2009).

shift toward towards defined contribution (DC) plans[2] and the recent economic and financial crises require an extensive amount of detailed data to provide pointed answers. Kröger et al. (2011) provide a comprehensive review of the data sources available for the analysis of old-age provision in Germany, documenting the shortcomings of each source. In general, the multiple trade-offs between detailedness and amount of information on the one hand side and precision and reliability of the answers on the other hand side strongly limit the analysis which can be carried out with data from a single source. A very promising way to improve both quality and quantity of the data is therefore the merging of information from different sources.

Thus, in the latest wave of the SAVE survey (run in 2011/2012) respondents were asked for permission to link their survey responses to administrative records stored at the Federal Employment Agency in order to supplement the information provided by the households. The linked data set offers new research avenues on, for instance, the link between employment histories and old age poverty or retirement decisions of couples.

The remainder of the article is structured as follows: In section 2 we give an outline of the data set, followed by a presentation of the linked data set in section 3. Information on data access is provided in section 4. Section 5 concludes.

2. Outline of the Data Set

Building up a panel of respondents who give meaningful answers on such sensitive and complex topics as saving behavior and wealth accumulation is not an easy task. In the following, we focus on the following aspects: questionnaire design, sampling scheme and interview mode, panel stability, as well as response to difficult items (item non-response).

Questionnaire design

In every wave a rather stable set of questions have been asked, designed such that the interview does not exceed 45 minutes on the average. This core questionnaire can be divided in six parts, briefly summarized in Table 1. From time to time, the core has been complemented with extra modules dealing with up-to-date topics. In 2009 and 2010, for example, specific questions were inserted to better measure and understand the effects of the economic and financial crisis on households' investment and saving decisions.

---

[2] DC plans are retirement plans in which the amount of the annual contribution paid in the plan is specified and invested in individual accounts. On the contrary, in defined benefits (DB) plans the monthly benefit on retirement is predetermined by a formula (usually based on earning history, age and tenure), irrespectively of investment returns.

Table 1

Structure of the SAVE questionnaire: core questions

Part 1:	Introduction; determining which person will be surveyed in the household
Part 2:	Basic socio-economic data of the household; health questions (since 2005)
Part 3:	Qualitative questions on saving behavior, income and wealth
Part 4:	Quantitative questions on income and wealth
Part 5:	Psychological and social determinants of saving behavior
Part 6:	Conclusion: comments on the interview

Sampling scheme and interview mode

SAVE consists of two main subsamples which differ in their sampling scheme: the *Random Sample* – drawn using a random sampling technique and the *Access Panel* – drawn from a standing panel of households surveyed at regular intervals, operated by the company TNS Infratest TPI.[3]

While the interview is conducted mainly as a self-administered paper & pencil questionnaire (P&P), the interview mode might differ between samples: A first and, if necessary, a second reminder are sent to those respondents who do not return the questionnaire within a certain time. Should the questionnaire be still missing after the second reminder, respondents in the Access Panel receive a third reminder by phone. In contrast, participants in the Random Sample are contacted by an interviewer to arrange an appointment. In those cases the questionnaire is eventually completed in a personal interview.[4]

Panel stability

After a first (mostly experimental) wave in 2001, a second wave was put into field in 2003/2004 and since 2005 the study has been repeated on an annual basis. In 2011/2012 the ninth wave has been collected. The panel has thus

---

[3] The first wave of SAVE was mainly conducted using a Quota Sample. After one re-interview in 2003 to obtain data on attrition rates, the Quota Sample was discontinued.

[4] Until 2009, the personal interviews were computer assisted (CAPI). Since 2010, they are paper assisted (PAPI). The sensitive questions on households' finances are however not part of the personal interview. They are left as P&P questionnaire dropped by the interviewer to be answered in private. See Essig/Winter (2009) for an analysis of different interview modes on answers to sensitive questions.

reached maturity and it is in a stage where the real value added of a panel survey can be harvested (cfr. Wagner et al., 2007).

Table 2 displays the development over time of the various sub-samples. The panel has reached a high stability: since 2008, the response rate from wave to wave has been around 90%. By 2010, about 55% of the Random Sample entering the panel in 2005 and 48% of the Access Panel interviewed in 2006 were still in the sample.

Table 2

Sample design and sample size over time

	Quota Sample	Access Panel 2001	Random Sample 2003	Random Sample: refresher 2005	Access Panel: refresher 2006
2001	1169	660			
2003/04	483	487	2184		
2005		357	646	1302	
2006		333	484	1021	1636
2007		321	431	902	1277
2008		276	393	848	1091
2009		234	352	765	871
2010		216	321	719	791

Source: SAVE 2001–2010.

Item non-response

Item non-response is a phenomenon that frequently occurs in survey data: for various reasons, respondents might prefer not to answer particular questions or items. For the vast majority of variables in SAVE, item non-response is not a problem, actually slightly better than missing rates documented in other surveys (Bover, 2004; Hoynes et al., 1998; Juster/Smith, 1997; Kalwij/van Soest, 2006). Typically, mainly due to privacy concerns and cognitive burden, there are higher rates of item non-response for detailed questions about household financial circumstances. For example for the amounts held in items such as stocks or bonds, nonresponse rates are between 11 and 17 percent of the households holding the items. The pattern is quite clear: the less defined the items are (such as "other assets" or "other debt") the higher is item non-response. While private old-age provision is reasonably well covered, households know very little about occupational pensions with resulting in high non-response rates (32% on annual balance).

Deleting the observations with missing values and relying only on a complete-case analysis not only reduces the sample size, but might also lead to

biased results, as item non-response may not be random among respondents. To handle this problem, missing values in SAVE are imputed using a multiple imputation procedure. This method aims at capturing all relevant relationships between variables in order to preserve their correlation structure. In a first step the data base's panel structure was used to logically impute several variables (Ziegelmeyer, 2009). Then the missing values in each single variable are imputed conditioning on as many relevant and available variables as possible (for further details, see Schunk, 2008; Ziegelmeyer, 2012). In the case of SAVE, the final user has five complete datasets, with all missing values replaced by imputed values. The differences in the imputed values across those five versions reflect the uncertainty about the "true" missing value. To get meaningful results, each of the completed dataset should be analyzed by standard methods, and the results should be combined to produce estimates and confidence intervals that incorporate missing data uncertainty (Rubin, 1996). In SAVE all the imputed values are however flagged, so users are free to ignore the imputations.

3. Linking SAVE to Administrative Records

An increasing number of large-scale social surveys around the world are now using some form of data linkage to supplement the information obtained from the survey with that included in administrative records. Since the advantages offered by these two kinds of data are strongly complementary, combining the two sources represents a very promising way to enhance the quality and the quantity of the information available to researchers (Calderwood/Lessof, 2009). To proceed with record linkage, however, respondents' explicit and informed consent is usually required (Lessof, 2009).

In 2011 respondents and their partners were asked for written consent to link their survey data with their administrative records stored at the Federal Employment Agency. The administrative records contain daily information on wages and social transfer payments received since 1975 for West Germany and, since 1991 for East Germany. Information on the participation in active labor market programs is included from the year 2000 onwards taken from the so called Integrated Employment Biographies (Dorner et al., 2010). The labor market histories are further enhanced with additional information on the employers at the establishment level (such as economic sector of activity or qualification and age structure of its labor force) drawn from the Establishment History Panel (Spengler, 2008). All individuals who have worked at least one day as an employee paying social security contributions in Germany are included in these administrative records.

SAVE is directed at the household level and hence, both the main respondent and the partner were asked for consent on separate forms. In 58% of the cases at least one household member agreed to data linkage. When considering all the individuals who have been asked for consent, we obtain a consent

rate of 49% as among the couples not all the partners agreed to the record linkage.

As the willingness to provide consent is not universal, a potential bias is introduced if individuals who consent systematically differ from those who do not. In a multivariate analysis, however, we find only mild evidence for consent bias. Apart from income and education, privacy concerns in combination with a general resistance towards the survey are found to be the main determinants of the consent decision.[5]

4. Data Access

The SAVE data are stored at the German Central Data Archive (ZA) in Cologne and are available upon a small fee for scientific research.[6] The use of the data for commercial purpose is explicitly not allowed. In addition, a compact dataset (miniSAVE) containing a selection of variables from the datasets 2005 to 2010 can be ordered via fax or e-mail directly from MEA. Due to its pre-aggregated variables the dataset is easier to handle and therefore especially suitable for teaching purposes.

Strict confidentiality rules apply for the use of the linked data set. On-site use at the Research Data Centre (FDZ) of the Federal Employment Agency and subsequent remote data access is possible. An application form has to be filled by the researcher, explaining the prevailing public interest of the data use and the scientific background. The application needs approval from the Federal Ministry of Labour and Social Affairs (BMAS). In that case a contract with the FDZ has to be signed. Further information on data access can be found on the website of the FDZ.[7]

By now more than 100 researchers have used the SAVE data for various projects, like the analysis of the role of subjective life expectancy on the demand for private old-age provision products (Doerr/Schulte, 2012), the effect of self-control and financial literacy on the ability to build up savings (Kaufmann, 2011), the analysis of the relevance of various saving-motives (Busl et al., 2012) or the relationship between the demand for Riester-products and that for unsubsidized private pension insurances (e.g. Pfarr/Schneider, 2012).

---

[5] For a first appraisal see Coppola and Lamla, 2012. Further results are available upon request.

[6] See http://www.gesis.org/en/services/data-analysis/data-archive-service/charges/.

[7] See www.fdz.iab.de.

5. Summary and Outlook

The link between institutions, saving behavior and old-age provision is complex and is still not well understood, even from a theoretical point of view. It becomes even more complex when taking into account the well-known behavioral and cognitive biases which affect individuals' decision making. The empirical investigation of this link, however, requires an extensive amount of detailed information which is almost impossible to collect using conventional household surveys. SAVE was constructed in order to combine detailed quantitative information on wealth and income with qualitative information explaining household saving behavior. In order to supplement survey responses and to improve data quality information from administrative records from the Federal Employment Agency has been merged. The linked data set offers new avenues for research regarding, for example, the relation between employment histories and old age provision and retirement decisions of couples.

However, there are still crucial gaps in the data, especially with respect to occupational pensions. Obtaining accurate information on these pension plans from respondents is almost impossible due to the formidable lack of knowledge displayed by respondents (Gustman et al., 2011). A possible solution to milder the lack of information is to extend the data linkage of the SAVE survey to a third source, namely an employer survey specifically focused on the occupational pensions.

6. References

Börsch-Supan, A. / *Coppola*, M. / *Essig*, L. / *Eymann*, A. / *Schunk*, D. (2009): The German SAVE Study: Design and Results. MEAStudies nᵒ5, Universität Mannheim.

Bover, O. (2004): The Spanish Survey of Household Finances (EFF): Description and Methods of the 2002 Wave, Documentos Ocasionales N.º 0409, Banco de España, Madrid.

Busl, C. / *Iliewa*, Z. / *Jokisch*, S. / *Kappler*, M. / *Roscher*, T. / *Schindler*, F. / *Schleer*, F. (2012): Sparen und Investieren vor dem Hintergrund des demografischen Wandels, Gutachten für das Bundesministerium der Finanzen, Berlin.

Calderwood, L. / *Lessof*, C. (2009): Enhancing Longitudinal Surveys by Linking to Administrative Data, in Methodology of Longitudinal Surveys (ed P. Lynn), Chichester, UK.

Coppola, M. / *Lamla*, B. (2012): Empirical Research on Households' Saving and Retirement Security: First Steps towards an Innovative Triple-Linked-Dataset, MEA Discussion Paper, 258 – 12, Max-Planck-Institute for Social Law and Social Policy, Munich.

Doerr, U. / *Schulte*, K. (2012): Betting on a Long Life – the Role of Subjective Life Expectancy in the Demand for Private Pension Insurance of German Households, Schmollers Jahrbuch 132 (2), 233 – 263.

Dorner, M. / *Heining*, J. / *Jacobebbinghaus*, P. / *Seth*, S. (2010): The Sample of Integrated Labour Market Biographies, Schmollers Jahrbuch 130 (4), 599–608.

Essig, L. / *Winter*, J. (2009): Item nonresponse to financial questions in household surveys: An experimental study of interviewer and mode effects, Fiscal Studies 30 (3/4), 367–390.

Gustman, A. L. / *Steinmeier*, T. L. / *Tabatabai*, N. (2011): The Effects of Changes in Women's Labor Market Attachment on Redistribution Under the Social Security Benefit Formula, Working Paper, 248, Michigan Retirement Research Center.

Hoynes, H. / *Hurd*, M. / *Chand*, H. (1998): Household Wealth of the Elderly under Alternative Imputation Procedures, inInquiries of Economics of Aging (ed. D Wise), Chicago, 229–257.

Juster, F. T. / *Smith*, J. P. (1997): Improving the quality of economic data: Lessons from the HRS and AHEAD, Journal of the American Statistical Association 92 (440), 1268–1278.

Kalwij, A. / *van Soest*, A. (2006): Item Non-Response and Alternative Imputation Procedures, in: A. Boersch-Supan / H. Juerges (eds.), The Survey of Health, Ageing and Retirement in Europe: Methodology, MEA Mannheim.

Kaufmann, C. (2011): The Influence of Information Presentation, Psychological Mechanisms, and Personal Characteristics on Households' Financial Decision Making, Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften der Universität Mannheim.

Kröger, K. / *Fachinger*, U. / *Himmelreicher*, R. (2011): Empirische Forschungsvorhaben zur Altersicherung – Einige kritische Anmerkungen zur aktuellen Datengrundlage, Working Paper, 170, Rat für Sozial- und Wirtschaftsdaten.

Lessof, C. (2009): Ethical Issues in Longitudinal Surveys, in: P. Lynn (ed.), Methodology of Longitudinal Surveys, Chichester, UK.

Pfarr, C. / *Schneider*, U. (2012): Choosing between subsidized and unsubsidized private pension schemes: evidence from German panel data, Journal of Pension Economics and Finance 12 (1), 62–91.

Rubin, D. B. (1996): Multiple Imputation After 18+ Years, Journal of the American Statistical Association 91 (434), 473–489.

Schunk, D. (2008): A Markov chain Monte Carlo algorithm for multiple imputation in large surveys, Advances in Statistical Analysis 92 (1), 101–114.

Spengler, A. (2008): The Establishment History Panel, Schmollers Jahrbuch 128 (3), 501–509.

Wagner, G. / *Frick*, J. / *Grabka*, M. (2007): The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements, Schmollers Jahrbuch 127, 139–169.

Ziegelmeyer, M. (2009): Documentation of the logical imputation using the panel structure of the 2003–2008 German SAVE Survey, MEA Discussion Paper 173–09, MEA Mannheim.

– (2012): Illuminate the unknown: evaluation of imputation procedures based on the SAVE survey, Advances in Statistical Analysis. 1–28. DOI: 10.1007/s10182–012–0197–2.