



MAX-PLANCK-INSTITUT FÜR SOZIALRECHT UND SOZIALPOLITIK
MAX PLANCK INSTITUTE FOR SOCIAL LAW AND SOCIAL POLICY

mea *Munich Center for the Economics of Aging*

**Empirical Research on Household's Saving and Retirement Security:
First Steps towards an Innovative Triple-Linked-Dataset**

Michela Coppola, Bettina Lamla

07-2012

MEA DISCUSSION PAPERS



Alte Nummerierung: 258-12

Empirical Research on Households' Saving and Retirement Security: First Steps towards an Innovative Triple-Linked-Dataset

Michela Coppola[§]

Bettina Lamla^{*}

This version: July 2012

Abstract: There is an increasing interest among social scientists in merging survey data with administrative records from social security institutions. Record linkage represents one way to combine different sources using a unique identifier such as the Social Security number. The informed consent of the respondents however is required, which in turn might induce bias to the consent question and even threaten stability in a panel study. Data from the longitudinal household survey "Saving and old-age Provision in Germany" (SAVE) are used for analysis of consent rates and patterns. In the latest wave of the study participants have been asked to provide their written consent to link their answers to administrative data from the Federal Employment Agency which also includes information on the respondents' employers. The combined data set will open new avenues for research on the link between institutions, saving behavior and old-age provision: The survey data contains information on private pension and non-pension wealth which will be complemented by complete employment histories. Moreover, from the administrative data entitlements to public pensions can be derived, while an employer survey will shed more light on the diffusion of occupational pensions. SAVE is mainly conducted as a self-administered paper and pencil (P&P) questionnaire, while existing research is based on personal interviews. Given a response rate of 81% of the participants and a consent rate of 58%, asking for consent appears feasible in a P&P design. There is evidence for mild consent bias. However, considering correlations between giving the consent and a series of socio-demographic variables, as well as variables capturing respondents' motivation and willingness can explain variation in the consent only to a small extent. We conclude that most of the variation is random.

JEL classification: C81; C83

Acknowledgment: We would like to thank Axel Börsch-Supan and Julie Korbmacher for their helpful comments, as well as Klaus Kortmann and Thorsten Heien for their valuable expertise which guided us during each phase of the fieldwork. The financial support of the German Science Foundation (DFG), which has generously financed the creation of the SAVE panel data set is also kindly acknowledged. Furthermore, we are grateful to Philipp Rösch for excellent research assistance.

[§] Munich Center for the Economics of Aging – Max-Planck-Institute for social law and social policy. Amalienstr. 33, D – 80799 Munich, Germany. e-mail: coppola@mea.mpisoc.mpg.de

^{*} Munich Center for the Economics of Aging – Max-Planck-Institute for social law and social policy. Amalienstr. 33, D – 80799 Munich, Germany. e-mail: lamla@mea.mpisoc.mpg.de

1. Motivation

The radical changes to the public pension systems in the last decades, the gradual shift of occupational pensions from defined-benefits (DB) towards defined-contribution (DC) plans as well as the recent economic and financial crises casted into doubt one of the mainstay of modern welfare states: old-age income security. When taking into account the well-known behavioral and cognitive biases which affect individuals' decision making the link between institutions, saving behavior and old-age provision becomes more complex. It is therefore unclear in which direction and to what extent these changes alter peoples' current behavior and future well-being.

A comprehensive analysis of individuals' and households' old-age security represents therefore a priority for both, social scientists and policy makers. Its empirical investigation requires an extensive amount of detailed information. The ideal data set should offer a reliable and comprehensive measure of wealth, covering not only public pension wealth, but also occupational pensions, private pension plans as well as private non-pension wealth. Furthermore, the shift towards DC plans implies increasing relevance of labor market biographies for the old-age income security while, at the same time, employment histories are becoming more fragmented and inhomogeneous. Therefore, the complete employment histories need to be included as well. Besides that, information on expectations, cognitive ability, and financial literacy should be provided to enrich standard economic models with behavioral elements. Finally, data with a longitudinal structure are required to correctly capture the effects of institutional changes and macroeconomic shocks. Since individual risks are usually pooled together and one partner may provide retirement security for the other, it is essential to collect all the information at the household level.

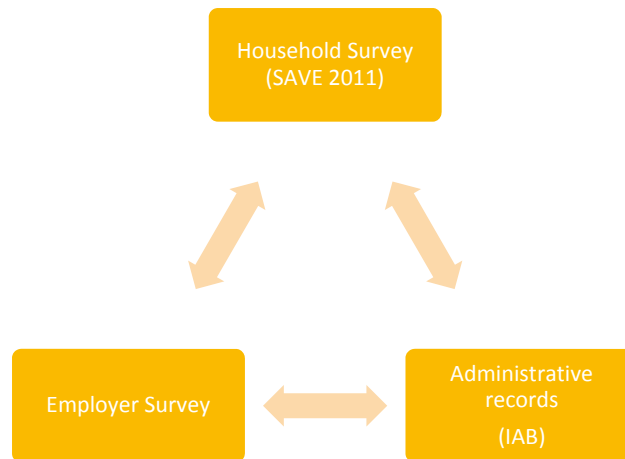
The construction of such an ideal dataset is clearly beyond the possibilities of conventional household surveys. In general, the multiple trade-offs between detailedness and amount of information on the one hand side and precision and reliability of the answers on the other hand side strongly limit the analysis which can be carried out with data from a single source. The complexity and the sensitivity of the topics related to the analysis of the old-age provision further sharpens these trade-offs. A very promising way to overcome at least part of the problems is therefore the merging of information from different sources. In particular, the combination of survey data (collected through questionnaires addressed to individuals, households, and/or firms) with administrative data (process-generated data produced by social security institutions while performing their institutional tasks) can produce a dataset which comes close to the ideal one described above.

"Record linkage" represents a possible way to merge survey data with administrative records: a unique identifier (like for example the Social Security number –SSN) is used to find exactly the same person in both data sources.¹ In the last years an increasing number of social surveys have been linked to a wide range of administrative data (e.g. the Health and Retirement Study in the US - Olson, 1999; The British Household Panel Survey in the UK – Sala et al. 2010). In Germany a similar attempt has been recently conducted within the Survey of Health, Ageing and Retirement in Europe (SHARE): survey data have been linked to data from the German Pension Fund (Korbmacher and Czaplicki,

¹ Survey data and administrative records can also be merged via "statistical matching". Rather than finding exactly the same person in both sources, statistical matching links cases which are statistically similar according to certain characteristics observable in both databases. See D'Orazio et al. 2001 for a short introduction.

2011).² However crucial gaps remain in the available databases. Most noticeably, it is missing at the moment a database which accounts for the large diversity of pension plans and which allows a precise reconstruction of the diffusion of the occupational pension among the population.

Figure 1: The triple-linkage strategy



To overcome these gaps, the latest wave of the study “Saving and old-Age Provision” (Sparen und Altersvorsorge – SAVE) has been used to test the feasibility of an innovative triple-linkage strategy. The basic idea is to merge through a record linkage procedure: *i*) household survey data with *ii*) administrative records provided by the Federal Employment Agency and *iii*) employer survey data (Fig.1). The project is conducted in close cooperation with the Institute for Employment Research (IAB), the research institute of the Federal Employment Agency.

SAVE represents an ideal database to test this strategy given its focus on households’ saving and old-age provisions coupled with its richness of relevant psychological and behavioral indicators and with its longitudinal structure (see section 2 for further details on the study). While the household survey provides detailed quantitative and qualitative information on household net worth, health status, expectations and financial literacy of the respondents, the administrative records provide the complete employment biography of the respondents and their partner. Using information in the administrative records it is also possible to identify the current employer. In a next step, the employers can be surveyed about types and rules of the pension plans offered at the establishment level.³

Several steps are necessary to build up the triple-linked dataset: first, SAVE respondents have to be asked for their SSN and their written consent to data linkage. Second, the IAB has to check whether the SSN is available and to extract the relevant records, which include also information on the current employer. Finally, the employers are contacted and surveyed.

² The German Socio-Economic Panel (GSOEP) has been also linked to administrative records via statistical matching (Rasner et al., 2011).

³ Ideally one would ask the employer directly about the pension entitlements of the respondent. That could be however too burdensome for the employer, who (in all the cases where the occupational pensions are outsourced) may not even have the required information. We therefore prefer to ask about the general rules governing the occupational plans in the establishment.

This paper documents the results of the first step of this procedure, namely the request of consent to data linkage. That is a tricky issue: willingness to give consent is not universal, thus reducing the number of linked records and potentially introducing bias (Jenkins et al., 2006; Sala et al., 2010). Furthermore, the request for consent might induce individuals to drop out of the study, thus compromising its longitudinal structure. Moreover, as SAVE is a self-administered survey, the absence of an interviewer - who can further motivate the respondents and answer their questions - casted doubts if data linkage was feasible at all.

The article is structured as follows: section 2 shortly introduces the main features of the SAVE study; the field work of the wave 2011 is described in section 3; section 4 then presents the results: the size of the response and consent rate, a first judgment about the quality of the collected information as well as the analysis of possible consent patterns; section 5 summarizes and concludes.

2. The SAVE Study

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 SAVE study 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 and 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.⁴

After the first (mostly experimental) wave in 2001, a second wave was put into field in 2003/2004. Since 2005 the study is repeated on an annual basis: in 2011/2012 the ninth wave of SAVE was collected. 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.

⁴ For a detailed description of the scientific background of the study and of its design see Börsch-Supan et al. (2009).

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

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.⁵ Figure 2 plots 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.

Starting in 2005, a set of incentives was introduced to motivate the participation in the study. Since 2007 those incentives have been given in advance, irrespectively of the subsequent participation in the study. As of 2010, participants in the Random Sample received 20€ cash while participants in the Access Panel received a present worth 10€. Between the waves respondents receive Christmas and Easter greeting cards and a short information brochure about the scientific results of the study as measure of panel care. The high response rates and the panel stability achieved in the last waves confirm the effectiveness of these strategies.

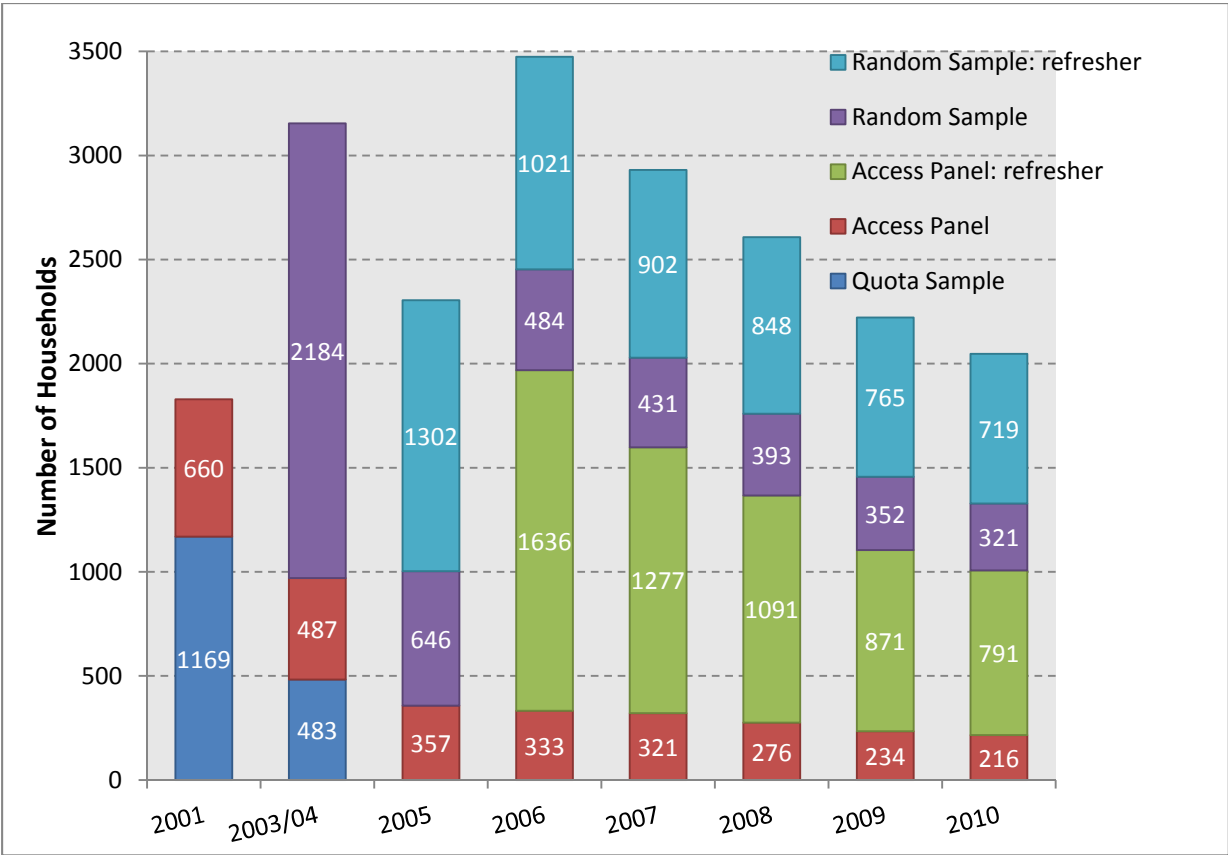
SAVE is conducted mainly as a self-administered Paper & Pencil questionnaire (P&P). A first and, if necessary, a second reminder are sent to those respondents who do not return the questionnaire within a certain time. Should the questionnaire be still missing after the second reminder, respondents in 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.⁶ The survey is usually fielded around Easter time. Spring is in fact a suitable time for asking questions about households' finances as this is the time when the annual notifications about the status of the financial contracts are sent out. That way it is easier for the respondents to retrieve the relevant information to answer the questionnaire.

⁵ 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.

⁶ 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 and Winter (2009) for an analysis of different interview modes on answers to sensitive questions.

Although the bulk of the questionnaires are returned within few weeks after the first mailing, due to the long series of reminders the field work generally stretches until late summer.

Figure 2: SAVE Sample design and sample size over time



Source: SAVE 2001-2010

3. SAVE 2011: Design and progression of the project

The ninth wave of the SAVE study was used to test the feasibility of the triple-linkage-strategy presented in section 1. It is conducted in close cooperation with the IAB, the research center of the Federal Employment Agency. Respondents were asked to give their written consent to linkage of their survey answers to their administrative records stored at the Federal Employment Agency. In case the respondent had a partner, his/her written consent was asked as well. More specifically, together with the questionnaire and the usual information material, the respondents received a form in which they were asked to report their social security number (SSN) or, failing that, their name, surname, sex, date of birth and current address. The respondents were instructed to fill in and sign the form and to send it back to the IAB using a prepaid envelope. An identical form (but in a different color) was provided for the informed consent of the partner.

A short reminder about the relevance of data linkage was inserted within the questionnaire, followed by the explicit request for filling the consent form. The placement of a question within a survey is extremely relevant: It is known for a long time that the context within which a question is asked impacts on respondent’s willingness to answer and/or on the kind of answer they give (e.g. Schwarz, 1991; Tourangeau and Smith, 1996). In the specific case of questions on consent to data linkage, it has been shown that asking for consent in the relevant context elicited higher consent rates than in

the case the question was asked at the end of the interview (Killpack and Oskala, 2011). In SAVE 2011, the request for consent was placed right after a set of questions on old-age income and before a battery of questions on future entitlements to public pensions. That is because we motivated the request of informed consent with the need to collect data to better estimate the future pension needs. Thus, questions concerning old-age income and entitlements from public pensions are the most salient from the point of view of the respondents.

The questionnaire itself has been considerably shortened (the total number of questions being reduced from 109 in the wave 2010 to 38) and focuses almost exclusively on old-age provision. After asking for the stable set of basic socio-demographic information, the questionnaire turns to the three pillars of the pension system asking the respondents about the ownership of occupational and private pensions as well as non-pension wealth, and about entitlements to public pensions. Most importantly, the whole set of quantitative questions on household's finances (part 4 of the core questionnaire) has been dropped.⁷ The reduction in the workload justified a reduction in the incentives paid out to the respondents: in both sub-samples incentives were halved.⁸

The fieldwork started considerably later than in the previous years, partially due to the challenges associated with the new design: In spring 2011 the respondents received a short brochure informing them about the delay in the process. The questionnaire was first mailed mid-November 2011. The first reminder was sent out as usual two weeks after the first mailing of the questionnaire, thus coinciding with the beginning of Christmas time. We postponed the mailing of the second reminder until January 2012 due to the Christmas holidays. The last phase of the fieldwork started in February 2012 and lasted until mid-March: about 200 households from the Random Sample have been contacted by an interviewer in order to arrange an appointment for a personal interview, while the remaining cases from the Access Panel have been reminded a third time via telephone.

A total of 1660 valid interviews could be obtained, representing 81% of the households giving an interview in the wave 2010. The response rate is clearly below the ones reached in the last waves, but it is still high in international comparison.⁹ In conclusion, neither the changes in the design nor the request for consent to data linkage significantly damaged the panel. That can be partially confirmed when looking at the comments the respondents left at the end of the questionnaire (Part 6 of the questionnaire, see Table 1). About 47% of the respondents used the opportunity to leave a general comment about the questionnaire. In 69% of the cases the comment is positive: among them, the shortening of the questionnaire and the dropping of the sensitive questions are praised in more than one fourth of cases. At the same time, an explicit complaint about the request of consent to data linkage is made in only 7 cases.

⁷ The entire questionnaire is available on line: www.mea.mpisoc.mpg.de

⁸ For the respondents in the Random Sample there was a further change in the design: due to legal reasons the incentives were sent in the form of cheques rather than cash.

⁹ The panel response rate for the Italian Survey of Household Income and Wealth conducted by the Bank of Italy has been relatively stable at around 75% (see Banca d'Italia, 2002- 2008). Only in 2010 it reached a peak at 82% (Banca d'Italia, 2010). The Spanish Survey of Household Finances run by the Bank of Spain reached a retention rate of 50% and 66% between the wave 2002 - 2005 and 2005 – 2008 respectively (see Bank of Spain 2005-2011).

4. Consent to data linkage

4.1. Consent rate and quality of the information

The IAB in Nuremberg received almost 1700 consent forms, coming from about 1100 households. Not all the received forms represent consent to data linkage: in 82 cases the form was left blank, indicating thus a refusal to data linkage. Table 2 shows the distribution of the number of consents given. In almost 600 cases only a single consent was given, while in more than 400 cases (representing about 63% of the couples in the sample) two consents were accorded. Altogether, 1008 households (totalling 1456 individuals) agreed to the data linkage.

However, not all the consenters took part in the wave 2011. In particular, for 51 households we obtained a written consent but the household questionnaire is missing. Although the answers of these households to older waves of the SAVE study will be linked to the social data, we exclude them from the analysis of the consent rate (see section 4.2). Out of the 1660 households who participated in the latest wave of the SAVE study, 956 gave an explicit consent to data linkage, representing 57.6% of the entire sample 2011.

Table 2: Number of households sending back the consent form by number of given consents

Consents	Frequency	Percent
0	82	7.5%
1	560	51.4%
2	448	41.1%
Total (households)	1,090	100%

Source: Consent forms, SAVE 2011.

Table 3: Consent status by participation status in the wave 2011.

	Frequency	Percent
Attriters, empty form	43	2.4%
Attriters, consent	51	3.0%
Panelists, no consent	704	40.1%
Panelists, consent	956	54.5%
Total (households)	1,754	100%

Source: Consent forms, SAVE 2011.

The consent rate is lower than that achieved in other studies. For example, in the wave 1992 of the US Health and Retirement Study (HRS) about 75% of the respondents gave their signed agreement to link their data with administrative records of the Social Security Administration (Olson, 1999); in the British ISMIE survey (a follow-up of the British Household Panel Survey – ECHP run in 2003) about 75% of the respondents gave the written permission to match their answers with administrative records about their benefits and tax credits (Jenkins et al., 2006); 77% of the participants in wave2 of

the English Longitudinal Study of Aging (ELSA) agreed to a linkage with administrative records on National Insurance contributions, benefits, tax records and pensions in 2004-2005 (Gray, 2010); finally, 73% of the participants in the third wave of the German part of SHARE gave their oral consent to a linkage with records from the German Pension Fund (Korbmacher and Czaplicki, 2011).¹⁰ Nevertheless, in studies with a focus on sensitive financial topics like in SAVE the consent rate is usually lower. The UK Family Resources Survey, for example, reached in the wave 2009/2010 a consent rate of 63% (McKay, 2012). In general, however, the consent rates cannot be directly compared with those achieved in SAVE, as the latter study has a self-administered P&P design while the above mentioned surveys are conducted as personal interviews.

The SAVE data are linked with the data from the Federal Employment Agency using the SSN as a unique identifier. Thus, the consent form asked the respondents (and their partner) to report their SSN or alternatively a set of information (e.g. name, date of birth, address) which are necessary to reconstruct the individuals' SSN. The quality of the information provided is good: in only 2% of the cases there are severe deficiencies in the reported data (wrong or missing SSN associated with incomplete criteria for reconstruction). In all the other cases, either the correct SSN (66% of the cases) or the complete set of criteria (32% of the cases) has been reported.

Table 4: Quality of the information to perform data linkage

	Frequency	Percent
Correct SSN provided	961	65.7%
Wrong SSN, full information	34	2.3%
Wrong SSN, incomplete information	3	0.2%
SSN not provided, full information	435	29.7%
SSN not provided, incomplete information	29	1.9%
Total (individuals)	1462	100%

Source: IAB communication

4.2. Pattern of consent

Finally, we look at possible selectivity in the pattern of consent, regressing the dichotomous variable “consent” – which equals 1 if at least one member of the household sent back a valid consent form, on a set of predictors. Besides controlling for a series of socio-demographic characteristics¹¹ (age, gender of the respondent, household composition¹², location in East Germany, retired status¹³, secondary and post-secondary education of the respondent, household net income quintile) another set of variables has been used to capture respondents' differences in ability and willingness to provide consent. The theoretical framework explaining response behaviour guides the choice of those indicators.

¹⁰ The required written consent, however, was collected only for 63% of the oral consenters.
¹¹ Data from the wave 2010 have been used to construct the socio-demographic controls used in the regression.
¹² Household's composition is controlled for using two variables: a dichotomous indicator equaling 1 if the respondent lives with a partner, and the number of children living at home.
¹³ This binary variable equals 1 if the respondent or his/her partner is actually retired.

Once asked for their consent to data linkage and their SSN, respondents have to proceed through four cognitive steps: *i)* understand and interpret the question; *ii)* retrieve any relevant information; *iii)* process and integrate the information into a summary judgment; *iv)* format this judgment into an answer (Tourangeau et al., 2000). Three factors affect the cognitive process: *i)* the difficulty of the task; *ii)* respondent ability and *iii)* respondent motivation (Krosnick, 1991). The task of giving an informed consent is indeed quite difficult not only because the respondents need to figure out what exactly data linkage means and which are the possible consequences of their consent, but also because they are asked to report their SSN, which might not be familiar to everybody. We therefore expect respondents' ability and motivation to play an important role in determining the decision to consent, as both affect and/or counterbalance the costs involved in such a difficult task. Furthermore, as emphasized by Singer (2004), individuals' attitudes toward risk and privacy are also relevant in the context of consent questions.

The longitudinal structure of the SAVE study and the wide range of indicators collected over the years offers a rich set of variables related to these issues. In the wave 2009 a short 3-item test to measure individuals' cognitive abilities was introduced in the questionnaire as part of a special module to measure financial literacy. The test, suggested by Frederick (2005), is a good predictor of cognitive ability in general and compares very favourably with more complex personality tests. Furthermore, it correlates with risk aversion and time preferences. In particular, good test scores are associated with less risk aversion and more patience (Oechssler et al., 2008). Each of the three questions has an intuitive (but incorrect) answer that comes quickly to mind. The questions are not difficult and indeed it is easy to verify whether the answer is correct. That, however, requires overcoming the impulsive response. Based on these answers, we construct an index which is 0 (*low cognitive ability*) for those respondents with up to 1 correct answer, and is 1 for the respondents with 2 to 3 correct answers (*high cognitive ability*). High cognitive abilities should help individuals in understanding the question and the scope of the consent, thus leading to higher consent rates. We expect that to be particularly true in a P&P context without an interviewer who can help the respondents in understanding the question and answer their possible concerns. We therefore interact the indicator for cognitive abilities with an indicator which equals 1 if the interview was conducted in the presence of an interviewer.

We also use the information on health status to control for further factors which might affect the cognitive process of understanding and answering the consent question. In particular, we construct a binary variable which equals 1 if the reference person and/or his/her partner reported to suffer from long-lasting health problems in the survey 2010.

We further control for the number of missing answers on sensitive items in the previous survey in 2010.¹⁴ High item-nonresponse may indicate both, lack of interest in the survey and concern with privacy issues. We expect therefore respondents with high item-nonresponse to those questions to be less likely to give consent.

In addition, we introduce a dichotomous variable which takes value 1 if the household was receiving unemployment benefits provided by the Federal Employment Agency in the previous wave. The idea is that such individuals might be more concerned about the data linkage. Given that the IAB is the

¹⁴ Sensitive items are: household's monthly income, value of real and financial assets, type and amount of outstanding debts.

research institute of the employment agency, and the SAVE questionnaire collects (although not in the wave 2011) detailed quantitative information on households' asset and income, those individuals might fear that their benefits (which are means tested) might be curtailed if their answers to the questionnaire are disclosed to the Federal Employment Agency.

Finally, a dummy variable which equals 1 if the respondent belongs to the Access Panel captures possible differences in the consent patterns of the two subsamples.

Table 5: Propensity to consent to data linkage (probit regression)

Variable	Coefficient	Standard Error
Age <30	Ref.	
Age 30-39	0.278	0.212
Age 40-49	0.108	0.201
Age 50-59	0.463 **	0.200
Age 60+	0.638 ***	0.213
RP is male	0.190 ***	0.067
RP has a partner	-0.050	0.086
# Children at home	-0.035	0.042
East Germany	0.190 ***	0.078
Household is retired	0.015	0.106
Unemployment benefits (t-1)	-0.083	0.138
Hauptschule	Ref.	
Realschule	0.180 **	0.081
Abitur	0.293 ***	0.110
No post-secondary	Ref.	
Vocational Training	-0.043	0.122
University	-0.127	0.159
Q1	Ref.	
Q2	0.098	0.121
Q3	0.374 ***	0.135
Q4	0.102	0.138
Q5	0.189	0.147
Access Panel	-0.027	0.067
Personal Interview	0.043	0.215
Cognitive Ability	0.055 *	0.032
Personal int. * Cog. Ability	-0.214	0.168
Long-lasting health problems	0.120 *	0.070
# missings sensitive items	-0.009 ***	0.003
Constant	-0.666 ***	0.217
McFadden's R2	0.0551	
Observations	1,660	

Significance levels: *** p<0.01; ** p<0.05; * p<0.10

Source: SAVE 2010, 2011

Table 5 presents the results. Only few socio-demographic characteristics are significantly related to the probability of consenting to data linkage. In particular, older households are more likely to give their consent as well as households with a male reference person (RP) and households located in East Germany. Secondary education is also positively correlated with the probability of giving the consent,

while there are no significant differences across different post-secondary educational patterns. Household income is only moderately correlated with the probability to consent: only households in the middle of the income distribution appear to be significantly more likely to give their consent in comparison to low income households. Korbmacher and Schröder (2011) find a similar pattern when analysing the consent rate in SHARE. As expected, individuals with higher cognitive abilities are also more likely to give their consent, although the coefficient is statistically significant only at a 10% level. Furthermore, the interaction term with the personal interview is not statistically significant, indicating that in the presence of an interviewer other mechanisms rather than individuals' cognitive ability determine the consent. Households suffering long-lasting health problems are also more likely to give their consent which is in line with previous research on consent patterns (e.g. Sala et al., 2010; Jenkins et al., 2006). Finally, individuals' attitudes towards sharing sensitive information (here captured by the number of missing items to sensitive questions in the survey 2010) are significantly correlated with the probability of consenting: respondents with a high number of missing items to sensitive questions are less likely to consent to data linkage.

Altogether, the variables considered in the analysis explain only a small fraction of the total variance in the probability of giving consent (McFadden's $R^2 = 0.0551$)¹⁵, thus suggesting that most of the variability in the consent rates is indeed random.¹⁶

5. Summary and conclusions

The analysis of individuals' and households' old-age security represents a priority for both, social scientists and policy makers. The empirical investigation of this topic, however, requires an extensive amount of detailed information which is almost impossible to collect using conventional household surveys. Merging survey data with administrative records provided by social security institutions via record linkage represents thus a very promising way to gain additional information. An increasing number of social surveys have been successfully linked in the last years. There are however still crucial gaps in the data, especially with respect to occupational pensions. A possible solution to milder the lack of information is to extend the data linkage to a third source, namely an employer survey specifically focused on the occupational pensions. The latest wave of SAVE, a longitudinal study focused on households' saving behavior and old-age provision, has been used to test the feasibility of the triple-linkage strategy.

The first step to build up such a triple-linked dataset is to collect respondents' informed consent to data linkage. This is not a trivial issue as such a request introduces a source of potential bias and might have negative consequences on the longitudinal structure of the study. The present paper has reported the results of this first step. Besides documenting the temporal progression of the fieldwork, we provided first evidence on: *i*) the size of the response and the consent rate; *ii*) quality of the information on the SSN provided by the respondents; *iii*) consent patterns.

Despite the request for consent to data linkage, coupled with a series of major changes in the survey design, a relatively high retention rate (81%) could be achieved. Almost 60% of the households in the sample agreed to record linkage. The achieved consent rate is lower than in other studies although

¹⁵ Despite being slightly higher than the McFadden R^2 , the estimated R^2 in a linear probability model is still low ($R^2=0.0733$).

¹⁶ Of course, only observable characteristics can be included in the regression. There is thus the possibility that much of the variability is due to unobserved characteristics.

the results are not directly comparable due to the differences in the survey design. In particular, the absence of an interviewer partially explains the result.

The quality of the provided information for the linkage is very good. About 66% of the consenters reported a correct SSN, while in more than 30% of the cases the full set of information needed to recover the administrative records is provided. Only in 2% of the cases the information is severely deficient.

Finally, we find only mild evidence of bias in the group of the consenters. In particular, we find that older individuals, households living in East Germany and male respondents have a higher probability to give their consent. Secondary schooling is also related to the probability of agreeing to data linkage, with better educated individuals being more likely to consent. Individuals' motivation and cognitive abilities are correlated with the probability to give consent. Interestingly, however, cognitive abilities become insignificant when the interview is conducted with an interviewer rather than as a self-administrated questionnaire. Altogether, the set of variable used in the regression explain only a very small part of the variance in the probability of agreeing to data linkage, indicating that most of the variability is indeed random. The results have to be interpreted with caution as we simply compare the group of consenters to the group of non-consenters who did not drop out of the panel. However, if the decision to continue participation in the survey is driven by the same characteristics which affect the probability to consent the results of our probit model will be biased. Our further research will shed more light on the selection problem.

The results reported here have shown that asking for consent to data linkage is feasible in a P&P design. In comparison with personal-interviews, P&P is much cheaper. Survey costs are not a trivial aspect of the entire design: the higher costs of a personal-interview should be justified only by a clear improvement in the results (Couper, 2000). When asking for consent to data linkage, that seems not to be the case. Some caveats, however, apply. The SAVE panel is a mature panel, containing experienced respondents who have been subject already to many waves of the study and have established a trustful relationship with the institution carrying out. Results might differ for novice respondents. Only further research can shed light on this.

Reference List

- Banca d'Italia (2002) Supplements to the Statistical Bulletin. Methodological notes and statistical information. Italian Households Budgets in 2000. Year XII, No. 6.
- Banca d'Italia (2004) Supplements to the Statistical Bulletin. Methodological notes and statistical information. Italian Households Budgets in 2002. Year XIV, No. 12.
- Banca d'Italia (2006) Supplements to the Statistical Bulletin. Sample Surveys. Italian Households Budgets in 2004. Year XVI, No. 7.
- Banca d'Italia (2008) Supplements to the Statistical Bulletin. Sample Surveys. Household Income and Wealth in 2006. Year XVIII, No. 7.
- Banca d'Italia (2010) Supplements to the Statistical Bulletin. Sample Surveys. Household Income and Wealth in 2008. Year XX, No. 8.
- Bank of Spain (2005): Survey of Household Finances (EFF): Description, Methods and Preliminary Results. *Economic Bulletin*. 3 – 21.
- Bank of Spain (2008): Survey of Household Finances (EFF) 2005: Methods, Results and Changes between 2002 and 2005. *Economic Bulletin*. 3 – 34.
- Bank of Spain (2011): Survey of Household Finances (EFF) 2008: Methods, Results and Changes since 2005. *Economic Bulletin*. 91 – 123.
- Börsch-Supan, A., Coppola, M., Essig, L., Eymann, A. and Schunk, D. (2009): The German SAVE Study: Design and Results. MEASudies n°5. Universität Mannheim.
- Couper, M. (2000): Web Surveys: A Review of Issues and Approaches. *The Public Opinion Quarterly*. 64 (4), 464 – 494.
- D'Orazio, M., Di Zio, M. and Scanu, M. (2001): Statistical Matching: a tool for integrating data in National Statistical Institutes. ETK-NTTS 2001 Conference: New Technique and Technologies for Statistics.
- 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.
- Frederick, S. (2005): Cognitive reflection and decision making. *Journal of Economic Perspectives* 19, 25–42.
- Gray, M. (2010): A Review of Data Linkage Procedures at NatCen. National Centre for Social Research.
- Jenkins, S., Cappellari, L., Lynn, P., Jäckle, A. and Sala, E. (2006): Patterns of Consent: Evidence from a General Household Survey. *Journal of the Royal Statistical Society. Series A.*, 169(4), 701-722.
- Killpack, C., Oskala, A. (2011): Understanding Society Innovation Panel Wave 4: Technical Report. National Centre for Social Research.
- Korbmacher, J., Czaplicki, C. (2011): SHARE-RV User guide. Available at: http://www.share-project.org/fileadmin/pdf_documentation/SHARE-RV/1_Methodenbericht.pdf
- Korbmacher, J., Schröder, M. (2011): Consent when linking survey data with administrative data. Fourth Conference of the European Survey Research Association (ESRA), Lausanne.
- Krosnick, J.A. (1991): Response Strategies for Copying with the Cognitive Demands of Attitude Measures in Surveys. *Applied Cognitive Psychology*. 5, 213 – 236.
- McKay, S. (2012): Evaluating approaches to Family Resources Survey data linkage. Department for Work and Pensions Working Paper no. 110.

- Oechssler, J., Roider, A. and Schmitz, P.W. (2009): Cognitive Abilities and Behavioral Biases. *Journal of Economic Behavior and Organization*. 72, 147 – 152.
- Olson, J.A. (1999): Linkages With Data From Social Security Administrative Records in the Health and Retirement Study. *Social Security Bulletin*, 62.
- Rasner, A., Frick, J. and Grabka, M. (2011): Extending the Empirical Basis for Wealth Inequality Research Using Statistical Matching of Administrative and Survey Data. *SOEPpapers 359*, DIW Berlin.
- Sala, E., Burton, J. and Knies, G. (forthcoming): Correlates of Obtaining Informed Consent: Respondent, Interview and Interviewer Characteristics. *Sociological Methods & Research*.
- Schwarz, N. (1991): In welcher Reihenfolge fragen? Kontexteffekte in Standardisierte Befragungen. ZUMA Arbeitsbericht 91/16. ZUMA-Mannheim.
- Singer, E. (2003): Exploring the meaning of consent: participation in research and beliefs about risks and benefits. *Journal of Official Statistics*, 19, 273-285.
- Tourangeau, R., Smith, T. (1996): Asking Sensitive Questions. The Impact of Data Collection Mode, Question Format, and Question Context. *Public Opinion Quarterly*, 60, 275 – 304.
- Tourangeau, R., Rips, L.J., and Rasinski, K. (2000) *The Psychology of Survey Response*. Cambridge and New York: Cambridge University Press.