



PARIS-JOURDAN
SCIENCES
ECONOMIQUES

HOW TO MEASURE RISK AND TIME PREFERENCES OF SAVERS ?

Luc ARRONDEL
CNRS-PSE

and

André MASSON
CNRS-EHESS-PSE

1 – INTRODUCTION

De gustibus and wealth inequality : how much may unobserved individual heterogeneity due to savers preferences – as stated by microeconomic theory – help to explain household wealth distribution and portfolio diversity in a cross-section ? The answer to that question has been the key objective of our long-lasting study based on two different data sets which combine information on households' socio-demographic characteristics and wealth on the one hand, and a specific set of questions allowing to infer individual preferences towards risk and time on the other.

The Insee survey "Patrimoine 1998" interviewed a representative sample of 10,207 households ; it was completed with a "pilot" Insee-Delta questionnaire of some ninety questions, aimed at measuring individual preferences, that was submitted in a second interview to a representative sub-sample of 1135 volunteers (see Arrondel *et al.*, 2004 and 2005 ; Arrondel and Masson, 2006).

The TNS-Sofres-Delta 2002 survey sent by mail to a representative sample of 2,500 individuals, in the 35 to 55 age bracket, a questionnaire integrating wealth data and information on preferences analogous to those gathered in the Insee survey, with additional variables such as religious education or political opinion. A similar questionnaire was independently addressed to 440 parents or children of the initial sample, in order to assess the degree of intergenerational "transmission" of preferences (see Arrondel and Masson, 2007, especially chapter 10).

In this paper, we shall focus on the methodological pitfalls raised by alternative ways of estimating individual preferences of savers. How to reach the optimal trade-offs between what the theory says about preferences and an operational way to assess the specific role of preferences in the observed diversity of wealth-related behaviours ? Or again, between parsimonious but largely inadequate models and more realistic but more complicated specifications ? Our contribution will indeed be mainly methodological, addressing the following basic issue : how to explain that our measures of preferences *perform so well*, compared to other studies, in explaining wealth accumulation and composition ? Moreover, how to validate our method of scoring that used a large number of questions to build synthetic and ordinal *scores* measuring, for each interviewed individual, such or such (parameter of) preference ? How to be (pretty) sure, for instance, that a *single* "risk score" is enough, that it is robust to other choices of questions, and that it really represents a summary indicator of attitudes and behaviours towards risk and not something else ?

The next section will further present the main features of the problem of measuring savers' preferences. The organization of the paper is at the end of this section.

2 – MEASURING SAVERS' PREFERENCES : METHODOLOGICAL ISSUES

Our objective is not, *per se*, the test of (structural) models of savings, as it is often the case in experimental work ; it is first to evaluate the explanatory power of preference heterogeneity on wealth inequality, while checking that the effects of a given indicator of preference are in agreement with theoretical predictions. What, then, the expression "as stated by microeconomic theory", in the beginning of the introduction, more specifically means ? The reference is *standard* life-cycle theory, augmented to bequests, which assumes full rationality and time consistency of saving behaviour, and more precisely : intertemporal additive preferences for consumption, with exponential discounting (DU model) ; maximization of expected utility under uncertainty (EU model) ; homothetic, that is isoelastic, instantaneous preferences ; and a separable bequest motive.

Under these assumptions, the theory is indeed parsimonious. Schematically, it characterizes preferences by only three parameters, with the following effects :

(1) a risk parameter (first, the degree of relative risk aversion) that affects precautionary savings and the share of risky assets – briefly called "prudence" ;

(2) a time preference for the present, or time discount rate, which determines the agent's behavioural horizon – the inverse will be called "foresight" ; it influences the age profile of consumption and saving for retirement, as well as the demand for long-term investments ;

(3) the degree of altruism towards children, which conditions the size and timing of intergenerational transfers (financial help, gifts and bequests), as well as the ownership of specific assets such as life insurance.

The model has indeed clear predictions. Total net worth, for instance, is expected to increase with "prudence" (owing to precautionary savings), "foresight" (life-cycle and retirement saving), and family altruism (planned bequests), the quantitative importance of the effects being quite sensitive to the value attributed to the parameters of preference. When investigating the role of preference heterogeneity, this standard theoretical framework will remain our benchmark ; but we shall have to accommodate the fact that it often leads to strongly *unrealistic* predictions, as claimed by the behavioural approach.

More explicitly, a lot of puzzles in wealth-related behaviours cannot be explained either by capital market imperfections, liquidity or borrowing constraints, or even by minor modifications of the standard model – intertemporal non separabilities of preferences (habits), tastes shifters over the life-cycle, utility of leisure or health, etc. –, but point out to imperfect rationality. These puzzles include the inadequacy of saving at retirement age of a substantial proportion of households, especially compared to resources available over the life-cycle (which should imply a sharp reduction in consumption at the end of life) ; the limited demand for shares, despite the difference in returns between stocks and shares (the "equity premium

puzzle"); the very low demand of pure life annuities, notwithstanding the fact that these products are especially designed to cope with lifetime uncertainty; the preference, as a disciplinary device, for "contractual", locked-in, saving products, even when they do not yield higher returns. For the behavioural approach at least, the explanation of such anomalies requires strong departures from the standard hypothesis of full rationality: loss aversion, insufficient propensity to plan for the future (Ameriks *et al.*, 2003), time inconsistency generated by hyperbolic discounting (Laibson, 1997), etc.

In this context, the behavioural approach and experimental economics seem to go hand by hand against the parsimony of the standard framework, in favour of a multiplication of preference parameters. The behavioural approach has carefully tracked any departure from full rationality, bringing to light many "bias", "anomalies", "errors" and the like. On the other hand, experimental studies tend to place the subjects in an artificial, *in vitro* situation, in order to avoid any framing effect and to reduce the impact of confounding factors; through somewhat abstract questions (choices between lotteries for instance) – like in experiments in physics –, they try to evaluate the exact value of a given parameter (say, relative risk aversion) or to test the validity of specific models or utility formalizations. Looking for models with more realistic predictions, these studies have then been led to introduce many new parameters, concerning as well behaviours towards risk or uncertainty – beyond the EU framework – as attitudes regarding time – against the DU specification (section 3).

When assessing the explanatory power of preference heterogeneity on wealth inequality or portfolio diversity, our contention is, however, that neither the experimental approach nor the behavioural approach would produce as promising results as our alternative way of measuring savers' preferences through our admittedly piecemeal *method of scoring*. This is mainly for practical reasons: the objective pursued requires that the number of preference parameters to be evaluated be kept as low as possible. The "hunt for a descriptive theory of choice under risk" (Starmer, 2000) or over time will introduce a dozen or more new parameters, the effect of which on the size or composition of wealth is often impossible to assess: what would be the effect of a higher aversion to ambiguity, or again a stronger taste for savouring on the amount of savings for retirement, for instance?

Our method of scoring will thus try to keep much of the parsimony of the standard theory while limiting the inevitable contributions borrowed from the behavioural approach. At the end, it proposes only four synthetic indicators of preferences for each interviewed individual, which concern a general *attitude towards risk* (rather than the standard parameter of relative risk aversion) and three preference parameters concerning time: *short-term*, mostly time-inconsistent *impatience*, *time preference* per se over the life-cycle; and family intergenerational *altruism*. We shall see that these indicators of preference have a richer content than in the standard framework; but the main innovation remains the addition of a

new parameter, short-term impatience, which may be partially connected to Laibson's parameter of hyperbolic discounting.

The method of scoring used appears, on the other hand, to take an opposite stand to the experimental approach, as described above. Instead of trying one or two types of questions that correspond closely to theory and focus on a specific domain (such as income or financial management), but are often too abstract (such as choices between lotteries), the idea is thus to multiply the number of more concrete, real-life questions, of different nature – opinions, intentions, behaviors, lotteries, hypothetical scenarios –, in different domains of life (consumption, leisure, work, health, financial investment and management, retirement, family and intergenerational transfers...), and to mix questions of anecdotic character with others, entailing long-term consequences or crucial issues. The large set of answers given by each individual is then summed up through synthetic, aggregate *scores* concerning the four preference parameters indicated above.

Our assumption is then twofold : first, we hope that the relevant questions will have a common component, closely related to the preference considered ; and second, that if it is actually so, the described procedure of averaging should reveal this component, while eliminating framing effects, confounding factors and the like. Now, this assumption is all the more warranted that our objective is limited : even though one may depict the resulting histogram of each score within the population (see graph 1a), these indicators of preferences are only *ordinal* measures ; we do not estimate a relative risk aversion of 3.6 for instance, as other, more ambitious methods of measurement would claim.

Section 4 underlines from the start the major drawbacks of this pragmatic approach : in short, our methodology of building scores is piecemeal, empiricist and moreover "agnostic", in the sense that we do not know which preference exactly we measure by our synthetic scores : for instance, is our risk score more an indicator of risk aversion, or of non-expected utility parameters – loss aversion, aversion to ambiguity... –, or again of irrational fears ? The basic issue raised in this methodological paper is then to understand why the scores *perform so well* – especially in comparison with alternative measures of preferences used in the literature – in explaining wealth accumulation and composition in the direction predicted by (standard) theory.

This issue breaks down in many different interrogations : How are framing effects and other confounding factors handled by the method of scoring ? Which type of "concrete" questions should be selected, in which domains on life, and how many of them – are the scores obtained enough robust to alternate choices in these matters ? Again, how to get "indirect" confirmations, at least, that the scores really measure the kind of preferences towards risk and time we are looking for, and not something else ? Finally, how to be pretty confident that so few scores are enough to assess the role of preference heterogeneity in household savings and investment.

Sections 5 and 6 deal with these issues, the former focusing on the procedure of aggregation used to derive scores from the answers given by the interviewees, the latter looking more for indirect validations of the scores as measures of risk and time preferences. Section 7 concludes while emphasizing a key lesson from our study : there is no easy shortcut, and it is more profitable to use a large number of heterogeneous questions – each with its defects –, rather than to look for a reduced set of presumably "ideal" questions.

3 – THEORETICAL BACKGROUND

A brief theoretical account will help to remind the trade-off between (standard) parsimony and (non standard) realism regarding the specification of savers preferences.

3.1. The parsimonious standard model

At age s , the consumer maximises under certainty the utility derived from her consumption stream until death, $[C(s)...C(T)]$, where $C(t)$ represents overall consumption at time t , and T is the age at death. The DU model (additive utility with exponential accounting) combined with homothetic preferences implies then a time independent instantaneous utility function u at time t (constant "tastes") of an isoelastic form :

$$U^s [C(s)...C(T)] = \int_{t=s}^T \alpha(t) u[C(t)] dt = \int_{t=s}^T \alpha(t) \frac{C(t)^{1-\gamma}}{1-\gamma} dt ; \quad \gamma > 0, \quad \alpha(t) \geq 0. \quad (1)$$

with γ (minus) the elasticity of decreasing marginal utility u . Individuals have a preference for the present : the weight accorded to future utility, $\alpha(t)$, falls over time, and the discount factor, $\delta(t)$ is defined as (minus) its logarithmic derivative:

$$\delta(t) = -\frac{d\alpha(t)/dt}{\alpha(t)} \geq 0, \quad \text{where: } \alpha(t) = \exp(-\delta t) \text{ si } \delta(t) = \delta. \quad (2)$$

Equation (2) corresponds to exponential discounting. Under time consistency, δ may be a function of age t but not of the distance from the present ($t-s$).

One sole preference parameter, δ , governs choice over time : the higher is δ , the shorter is the decision horizon (for given T). And the elasticity γ represents also the inverse of the intertemporal elasticity of substitution, denoted σ : the aversion to variations of consumption between dates ($1/\sigma$) is the same as the aversion to variations of consumption between states of nature (γ).

Under uncertainty, the individual is assumed to maximise expected utility and γ is then equal to the degree of relative aversion to risk, ρ . More generally, *all* risky behaviours – precautionary saving generated by prudence (the third derivative of u), the

reduction of portfolio risk in the presence of a background risk over labour income, which depends on temperance (the fourth derivative of u) and so on – will depend only on the parameter γ . Hence, the key relation in the standard framework :

$$\gamma = \rho = 1/\sigma. \quad (3)$$

Augmented to (voluntary) gifts and bequests, standard theory adds in the simplest case a third parameter, the degree of family intergenerational altruism, say θ , measuring the relative weight attributed to children well-being, so that the set of relevant preferences for saver i comes down to the triple $(\gamma_i, \delta_i, \theta_i)$. A very parsimonious model indeed, all the more so if one acknowledges that, for given (financial) wealth, portfolio choices are mainly driven by a sole parameter, that is γ .

The standard framework has even more powerful predictions that can be most clearly seen in the no-bequest version. Effects of preferences are highly non linear ; more precisely, the model engenders four specific accumulation regimes according to the values taken by the *couple* of parameters (δ, γ) . A low δ and a high γ define the prudent and farsighted "Conservative investors" who follow the representative age-wealth pattern of the life-cycle with "hump plus precautionary saving" (Modigliani, 1986). A high δ and a high γ leads to the buffer-stock model of the "Short-sighted prudent", following a "target saving behaviour" (Caroll, 2004). And one can likewise define risk-tolerant and short-sighted "Hotheads" (high δ and low γ) and risk-tolerant and far-sighted "Enterprising" (low δ and low γ) with also distinct predictions. This leads to an operative typology of savers with different wealth-related behaviours.¹

3.2. Non standard models : a profusion of preference parameter for realism

Homothetic preferences is a maintained hypothesis of the standard framework that is not usually supported by the data. But relaxing this assumption is costly in terms of preference parameters. Even in the additive and expected utility specification, risk aversion has now to be distinguished from prudence, temperance, and the like : Pierre may be more prudent than Paul but at the same time less temperant, and that will have bearings in their comparative precautionary savings and demand for risky assets under a given (background) income risk.

When you reject the time additive-expected utility specification, equation (3) breaks down. Recursivity instead of separability is enough to disentangle $\gamma = \rho$ from $1/\sigma$ (Kreps and Porteus, 1978), and (adding) non-expected utility on static gambles requires to distinguish between γ (concavity or u) and ρ (relative risk aversion). To account for a number of "anomalies" found in experiments (such as the Allais paradox) or in insurance demand, two

¹ See Arrondel *et al.* (2005) and Arrondel and Masson (2007). Technically, the two frontiers separating the four groups of savers are $\gamma \geq 1$ and Caroll's (2001) impatience condition.

more extensions seem necessary. The first one introduces a non-linear transformation of probabilities, which accounts for experimental subjects' greater sensitivity to (extreme) low or high probabilities and classifies them as "optimists" or "pessimists" : it is often claimed that it specification requires two parameters in order to capture the diversity of individual behaviours (Starmer, 2000). The second extension is even more fundamental in that it assumes that investors derive utility not from wealth *per se* but from deviations from a given level of reference (such as the current level of wealth). Especially, loss-aversion, as described by Kahneman and Tversky (1979), allows to explain why some risky assets do not appear attractive : the low demand for life annuities (respectively life insurance) can be partly attributed to the aversion to the risk of a zero-return investment in case of premature (respectively late) death.

Suggested by experimental and other work, the rejection of the time additive-discounted utility model will introduce an additional set of preference parameters. Habit formation may sometimes look like time preference when they entail ratchet effects ; on the other hand, phenomena of anticipation, such as the pleasure of savouring a pleasant event in the future, or the dread of a future unpleasant experience, introduce a bias in favour of the future – a kind of "negative" time preference. A stronger departure from the DU model is hyperbolic discounting, with a higher discount rate for the near future than for distant events. This can be presented in discrete time as follows (Laibson, 1997) :

$$U^t(C_t, C_{t+1}, \dots, C_T) = u_t(C_t) + (1-\beta) \sum_{k=1}^{T-t} (1+\delta)^{-k} u_{t+k}(C_{t+k}), \quad \text{avec: } 0 \leq \beta \leq 1. \quad (4)$$

In addition to the long-term discount rate, δ , we have a second short-term discount rate, β , which leads to time-inconsistent choices. Our score or short-term impatience is designed in part to capture this parameter β .

These developments deliver two intertwined bad news. First, the loss of parsimony, leading to a sheer impossibility of measuring all preferences : a (more) realistic non-standard model "consumes" far too many relevant new parameters. Second, the difficulty of measuring independently one given parameter, owing to cross-pollution by other preference parameters and other confounding factors. There is a somewhat vain quest to find the right, ideal question that would identify the relevant parameter while getting rid of the effects of other factors.

This is especially the case with the rate of time depreciation δ : an individual may prefer "one today to two tomorrows" because she dislikes waiting (a high utility of leisure), or because tomorrow her ability to enjoy consumption will be smaller (falling needs), or again owing to an uncertain future ("A bird in the hand is worth two in the bush"). The difficulty of adequately controlling these factors – as well as "savouring" ore "dread", working in the opposite direction – explains in part the diversity of measures. But the enormous instability of δ -estimates in empirical studies, from – 6 % to 200 % (see Frederick *et al.*, 2002) – has still

other sources, such as too artificial questions and, especially, identification errors : δ is not to be confused with the interest rate (as in trade-offs between monetary gains or losses at different dates) nor with the intertemporal marginal rate of substitution for consumption. In other words, many failures of δ -measurement are first due to conceptual flaws. In this methodological paper, we shall tend to sidestep these conceptual issues and, therefore, to concentrate rather on the preferences towards risk.

4 – THE "SUCCESS" OF A HETERODOX MEASURE OF SAVERS PREFERENCES

With its emphasis of realism, the behavioural-non-standard approach seems to lead to a dead end : a proliferation of parameters of risk and time preferences, the effect of which on wealth-related behaviours is, moreover, barely known.

4.1. A middle, pragmatic approach

The opposite option is to focus on parsimony while using the standard framework as a maintained hypothesis. In the case of risk, the best illustration is the Barsky *et al.* (1997) well-known measure of relative risk aversion γ , based on lottery choices. Instead of her actual lifetime income, say R , the subject is offered various job contracts in the form of a lotteries, with one chance out of two to earn twice more and one chance out of two to earn only λR with λ a parameter inferior to one. In the standard framework assuming expected utility, the subject with indirect utility V will prefer the contract to the sure gain R only and only if :

$$1/2 V(2R) + 1/2 V(\lambda R) \geq V(R), \quad (5)$$

with V assumed to be isoelastic of parameter γ . By varying the value of λ , you can then determine a range of variation for relative risk aversion γ : for instance, if the subject refuse the job contract for $\lambda = 2/3$, but accepts it for $\lambda = 4/5$, the value of its parameter γ is in the interval $[2 ; 3.76[$. Then you can test if, everything being equal, the share of risky assets in total (or financial) wealth is proportional to $(1/\gamma)$ or, at least, is higher for subjects with lower values of γ ; or again, test if a higher γ leads to greater total (or financial) wealth.

This parsimonious methodology has serious shortcomings, however. First, as the authors acknowledge, it is *directly* based upon the standard model – "with as little departure from the theorist's concept of the parameter as possible" – despite the fact that this model leads to unrealistic predictions. Second, it relies on an abstract and complicated question concerning a very hypothetical situation – a question that subjects may have difficulty to grasp or feel poorly interested in : indeed, over ten percent refuse to respond, and a majority of the others reject all (four or six) risky contracts offered, choosing each time the sure

income option. Third, the choice concerns only the professional domain, and this begs the question of the heterogeneity of subject's reactions towards risk from one domain to another, such as health, family or consumption². Fourth, the answer given depends on the respondent's current situation, especially on the degree of risk of his/her portfolio³. Last, and more importantly, empirical results regarding the effects of the lottery choice measure on wealth-related behaviours are rather disappointing : on French data, for instance, the effect on total or financial wealth is not significant (Arrondel, 2002) ; and there is only a weak negative effect on the ownership of stocks and shares.⁴

Our idea was then to find a medium way between the realistic non-standard approach with too many parameters and its parsimonious alternative, relying on a maintained standard hypothesis. Taking the standard framework only as a loose benchmark, our methodology thus tries to compute a limited number of scores of preferences ; these scores are obtained through averaging the answers given to a large set of concrete, real-life questions of heterogeneous nature and covering many different life-domains. Previous empirical investigation led us to look for estimates of four parameters of preferences : risk attitude, time preference, altruism and short-term impatience. There are, however, two qualifications.

Our estimated scores have indeed *some* connections with the preference parameters of an "extended standard model" which would allow for voluntary bequests and hyperbolic or other inconsistent discounting, namely, in our notations :

$$(\gamma ; \delta ; \theta ; \beta) ; \tag{6}$$

and this model has clear predictions as to the effects of preferences on wealth-related behaviours. But our scores have some connections *only*, because our method of scoring does not attempt at all to ask hypothetical questions with "as little departure from the theorist concept as possible". And rightly so, in some ways, given that the model in question remains still unable to accommodate a number of puzzles regarding savers' behaviours.

Moreover, statistical analysis might as well reveal that the preference towards risk, for instance, has several heterogeneous dimensions so that two or more risk scores should be introduced, one concerning (say) small or anecdotic risks, another large risks with long-term or irreversible consequences, and so on. There is of course a pitfall : if too many different scores are required to capture, empirically, a given preference, our method loses much of its interest : in short, data will have the last word.

² Barsky *et al.* (1997, pp. 550-1) explicitly consider this point, which is called by psychologists the issue of "behavioural specificity".

³ A risk-tolerant may choose a risky portfolio and, consequently, be induced to choose a safe income stream, since she is already exposed to considerable risk : the Barsky *et al.* measure will then consider her, wrongly, as strongly risk-averse.

⁴ *Cf.* Arrondel and Calvo-Pardo (2002). For as discouraging results for the US, see Barsky *et al.* (1997), and for the Netherlands, Kapteyn and Teppa (2002).

4.2. Our method of scoring : a piecemeal, empiricist and agnostic methodology

Before going any further in the description of our method of scoring, it appears therefore important to state briefly, from the outset, the strong challenges that this methodology has to face ; that will make even more compelling the issue about how to interpret the "success" of the preference scores.

At a general level, our method of scoring raises three main problems.

(1) It follows a *piecemeal* approach. The choice of the questions is in part arbitrary. Their minimal number that would allow, through aggregate scoring, to get rid of framing effects is, ex ante, an open issue. Moreover, the selection of the life-domains to be covered is itself arbitrary – which domains, how many questions in each, for which preference ? – but also *contingent* to the type of behaviour to be explained. Since we want to explain wealth accumulation and composition, the questionnaire about preferences has only few and marginal questions directly related to saving and portfolio management ; on the other hand, if we were to study drug addiction for instance, questions in this financial domain might prove to be useful.

(2) Our approach can similarly be accused of being *empiricist*. As we have stressed earlier, it keeps as a loose benchmark an extended standard model (augmented to bequest and hyperbolic discounting) which cannot predict a bunch of wealth-related behaviours. Also, for each of the four preferences thus selected, the adequate number of scores to be introduced for remains a purely empirical issue : in other words, statistical analysis will alone tell if the global risk score – derived from the whole set of questions presumably related to risk behaviours – shows a sufficient degree of internal consistency, or if it is preferable, for instance, to introduce different "sub-scores" per domain of life.

(3) Finally, and more importantly, our method of scoring is quite *agnostic* about what the scores precisely measure. This ignorance may be a matter of degrees, given here in the order of decreasing importance in the case of risk or uncertainty : (i) is the risk score really an indicator of preferences or attitudes towards risk or uncertainty, or does it reflect something else – such as a kind of conservatism ? (ii) in the first (favourable) case, are these preferences or attitudes rational, or does the risk score represent rather irrational fears ? (iii) finally, to which specific preference parameters of the theory would be this risk score most associated : risk aversion, prudence, aversion to ambiguity, loss aversion, or again optimism/pessimism and other perceptions or beliefs regarding risk ?

There is no definite answer to objections (i), but only "indirect" validations of the scores indicating that the latter do mainly represent preferences towards risk or time (section 6). Concerning (ii), there is reasonable evidence that each score represents a *mixture* of rational and irrational preferences, the rational component being dominant for time preference

and altruism, and the irrational dimension most important for short-term impatience⁵. Finally, critique (iii) is less disturbing with only relative measures of preferences : as far as risk is concerned, for instance, the exact parameter measured is not a crucial issue as long as the departure from expected utility and homothetic preferences is limited : if Pierre is more risk-averse than Paul, then he should be also more prudent most of the times, etc. On the other hand, the fact that the risk score encapsulates also (distorted) perceptions of risk and similar beliefs is more problematic, since these elements are presumably much more volatile than genuine preferences.

4.3. The greater performances of our way of measuring individual preferences

In any case, these three general criticisms, which can be addressed from the start to our method of scoring, may appear so devastating as to make the reader quite sceptic about its usefulness. Why making such a fuss and devoting so much effort to implement an approach that lacks so obviously rigor and coherence ?

A first, easy, answer, is that alternative measures of individual preferences of savers have proved quite unsuccessful, a failure that has allowed debatable conclusions concerning the determinants of households saving behaviour. Thus, Venti and Wise (2001) divide the factors explaining wealth between "chance" factors, due to circumstances that are imposed to the agent, and "choice" factors, which they identify with the contribution of preferences. Then, they use regressions to evaluate the (unsurprisingly) limited explanatory power of variables assumed to represent circumstances and attribute the rest, rather casually, to difference in preferences, concluding that the "bulk of the dispersion of wealth must be attributed to differences in the amounts the households *choose* to save", that is ultimately to differences in preferences. Likewise, Americks *et al.* (2003) take advantage of the modest explanatory power of wealth inequality obtained by shaky measures of "standard" preferences toward risk and time – such as Barsky *et al.* (1997) measure for risk – to advocate their measure of the "propensity to plan", derived from a behavioural approach ; yet, the effect on wealth on this preference is not of great quantitative importance, and the test of endogeneity relies on rather weak instruments.

By comparison, it is useful to emphasize the so-called "success" of our scores of preferences with respect to the objective pursued. Remember that this objective does not concern the test of a structural model, but focuses on the explanatory power of scores on the amount and composition of wealth, with effects going in the direction predicted by

⁵ For instance, our score of time preference has a sizeable negative effect on net worth and long-term saving, as predicted by the standard model ; but the fact that the probability of inadequate savings at the eve of retirement increases significantly with the same score is evidence that the latter is also an indicator of a time-inconsistent lack of foresight or self-control.

theory – whether standard or not⁶. For that matter, we shall therefore adopt an unusual reverse order of exposition, while stressing first the final results obtained in our study before investigating further the methodology used to build the scores (section 5) and assessing the reliability that can be attributed to these measures of preferences (section 6).

The final success of our method can then be judged on two different grounds :

(i) the *quality* itself of the scores as measures of a given preference towards risk or time, according to various criteria made more explicit below, such as parsimony, a "well-distributed" histogram, or individual observable determinants – given a priori thoughts about their effects ;

(ii) the *explanatory power* of scores on wealth-related behaviours, and more precisely the significance, direction (as stated by theory) and quantitative importance of their effects on wealth characteristics, while taking into account endogeneity issues.

Moreover, the evaluation will be made partly in relative terms, by comparison with other measures of preferences proposed in the literature :

- *scales*, graduated from 0 to 10, according to the perception that the interviewee herself has of her attitude with respect to risk – between "prudent" and "risk-loving" –, preference for the present over the life-cycle – between "live for today" and "take care of the future" –, and short-term impatience – between "impatient" and "sedate" (Insee data) ;

- theory-based specific questions, such as Barsky *et al.* measure of risk tolerance on income lotteries or, regarding rational or irrational time preference, the ability or "propensity to plan" or, in our surveys, the existence and length of ground projects or plans : "do the interviewee has (or had) made plans in a number of different areas of their life (career, family leisure time, wealth etc.) over a period of 10, 20, or 30 years or more ?".⁷

Regarding (i), the quality of our measures, our first and main success is empirical parsimony : as we shall see in the next section, four synthetic scores appear "enough" from a statistical point of view – exhibit sufficient internal consistency – to capture, in different areas of life, the vast array of behaviours towards risk or related to time preference, short-term impatience and altruism. Moreover, each score has a minimal dispersion and, more precisely, a "nice" histogram, as shown in the case of risk on graph 1a on Insee data. From the "points" (– 1 ; 0 ; + 1) given to each question concerning risk (see below), maximum risk tolerance would amount to – 46, and minimum to + 45 ; the final risk score has a quite dispersed and regular distribution between – 22 and + 27, with mean 4.2, first quartile up to 0 approximately, and last one from 9 onwards. Note the strong similarities with the self-positioned (general) risk scale, from 0 ("prudent") to 10 ("risk-loving") – where the axe of

⁶ At the level of generality considered, most non-standard models will thus predict, as well as standard ones, that wealth should increase with "foresight", "prudence" and altruism, and that the ownership or share of risky assets should increase with any measure of preference linked to risk tolerance.

⁷ The variable takes thus four values : no long-term plans, whatsoever ; plans over 10 years (but less than 20) ; plans over 20 years (but less than 30) ; plans covering 30 years or more.

risk tolerance is reversed –, depicted in graph 1b ; in the latter case, however, there is a problem of anchoring, the value 5, in the middle of the interval, being the most often chosen. On the other hand, other measures of risk tolerance show much less dispersion : the Barsky *et al.* measure takes thus four values, and in most studies more than half of the subjects are in the less risk-tolerant group.

Another, more indirect, way to judge the quality of preference measures is to see how the latter can be explained by respondent's or household's characteristics. Whenever the dependent variable is highly subjective, the explanatory power of observables is generally weak, and this is also the case here. However, the explanatory power of the scores regressions is always much higher than the one of scales regressions, and moreover, very low for measures derived from theory-based specific questions⁸. Also, effects on scores "make sense", given what we know beforehand from theory and previous empirical studies : thus, all measures of risk attitude in our surveys say that men have a higher propensity to take risks than women, and that "prudence" increases with age, but effects are much stronger for the risk scores ; likewise only the time preference score tells that being married, having children, or the level of education exerts a significant negative impact on time preference, precisely as expected by Fisher (1930).

Other and more involved comparisons are also possible, regarding for instance the *transmission* of preferences. The role of social origin or religious education is more significant in scores regression, but still limited. On the other hand, the intergenerational correlation of preferences – between, for instance, father's or mother's time preference and child's time preference –, that can be evaluated in the TNS-Sofres survey are quite high when measured by corresponding scores (0.23 for risk, 0.15 for time preference, 0.12 for altruism), in the range of those for income, revealing a highly specific cultural transmission ; the same correlations are found insignificant for specific questions like Barsky *et al.* measure or the propensity to plan.⁹

Regarding now (ii), the (econometric) effects on alternative measures of preferences on wealth inequality and portfolio diversity, the superiority of scores is striking. If you look at differences between the first and the last quartiles of scores, net worth (e.g.) is thus found, as predicted by theory, to increase by 50 % with "prudence", 80 % with foresight, and 35 % or so with altruism – short-term impatience having no significant effect (see Graph 2). And altogether, the joint contribution of the three significant scores of preferences on wealth inequality is around 10 % for the total population and 17 % for the wage-earning sub-population : it is lower than the one of reference variables (age, income, social class or inheritance) but generally higher than the one of other variables : education, marital status, social origin, town size, etc. By contrast, among the scales, only the one of time preference

⁸ See Arrondel and Masson (2007), chapter 6 for the attitude towards risk, and chapter 9 for time preference, short-term impatience, and family altruism.

⁹ See Arrondel and Masson (2007), chapter 10.

has a significant (but smaller) effect on wealth. The Barsky *et al.* has no effect at all, and the propensity to plan only a weak positive effect.

The above discussion ignores the question of reverse causality between preferences and wealth : the rich may take more risks, have more foresight, or again be more altruistic with respect to their children. Tests for the endogeneity of our preference scores – using rather weak instruments that nevertheless pass the standard quality and validity tests – appear successful. This is not always the case of alternative measures : in particular, the planning variable is endogenous, for gross wealth at least.¹⁰

Similar conclusions in favour of our scores can be drawn concerning portfolio composition. Thus, the statement of Barsky *et al.* (1997, p. 560), that "yet, the relationship between [their measure of] risk tolerance and the holding of risky assets is much weaker than theory suggests it should be" applies as well to other measures of risk preference with the exception of our risk score.

Incidentally, let us note that, as predicted by theory, the couple of the risk and time preference scores appears an even better predictor of saving and, especially, of asset demands. Thus, knowing that an individual is risk-tolerant, or that she is short-sighted brings only a limited amount of information regarding wealth-related choices. But knowing that she is both at the same time – risk-tolerant and short-sighted – is far more useful regarding her saving and investment decisions : we know that she follows a buffer-stock behaviour, with target saving (§ 3.1).

So far for the final success of our scores, especially when compared to the rather disappointing results of other measures of preferences. How can we explain it ?

5 – THE METHOD OF SCORING DISSECTED

Remember that the Insee or TNS-Sofres questionnaires designed to evaluate preferences towards risk and time cover a wide variety of life-domains (consumption, leisure, health, financial lotteries, work, retirement, and family) and, for each of these domains, ask different types of questions (concerning behaviour, opinion, or choices over lotteries or fictitious scenarios).

5.1. Four problematic steps

How have these questions been chosen ? How have summary scores been derived from the answers given by the respondent ? We follow a four-step procedure, each step encountering serious difficulties.

¹⁰ For a discussion of endogeneity issues, and the view of scores as a collection of "natural" instruments, see below, § 5.3. and especially the Appendix.

(1) The initial step is, obviously, the *design* or formulation of the different questions. The main preoccupation is to vary the type and context of mostly concrete, real-life questions, while trying to limit *ex ante* the impact of unavoidable framing effects : presumably, each question should first concern preferences towards risk or time, and ideally reveal only one of the four parameters retained... This is the most tricky stage of our method of scoring, since the relevance of the choices made can only be judged afterwards, while looking at the performances of the scores derived or at their sensitivity to alternative choices or data sets. Questions allowing for too many possible interpretations or involving too many uncontrollable qualifications should be abandoned. To give one example or rejected questions : in retirement saving products or in natural experiments, the choice between a one-time lump-sum payment or an annuity payment is supposed to reveal time preference (see Frederick *et al.*, 2002, p. 385) ; but this decision involves far too many other confounding factors – liquidity constraints, loss aversion, transmission motive, life expectancy, whether the annuity is price indexed, etc.

(2) Granted that the question considered reveals "something" about the respondent's preferences regarding saving, the next step is to decide which preference, i.e. the *affectation* of the question to one – or more... – of the four components selected : risk tolerance (in the general sense), time-preference, short-term impatience, and altruism. The design of each question should help to make these *a priori* methodological decisions not too arbitrary. Yet, there is no way to get rid of *double meaning* problems, especially since the future may be depreciated either because it is uncertain or because of time discounting (a pure priority given to the present), or for both reasons at the same time. Let us give examples to be more specific.

A number of questions may be reasonably assumed to concern only risk behaviours. Anecdotal ones, such as : "do you take an umbrella when the weather outlook is uncertain ?" ; or "do you park your car where you would run the risk of a fine ?". Lottery choices. Consumption habits such as : "have you changed your consumption patterns as a result of the Mad Cow disease ?" (remember that the surveys were carried out in 1998 and 2002). Opinions : "do you agree with the statement "marriage is a form of insurance ?" ; or "do you feel concerned by current public health issues (such as AIDS or contaminated blood) ?".

Some questions have been especially designed as to elicit pure time preference, such as : "as a result of an unexpected workload, your employer asks you to forego a week's holiday this year, in exchange for x extra days next year... : do you agree ?". Presumably, other reflect only short-term impatience (to stop reading a book if the first few page are not to your liking, to lose often one's patience when queuing up).

On the other hand, risk and time (preference) elements are indissociably intertwined in the following questions that appear to have significant contributions to both corresponding scores : "do you think that it is worthwhile to gain a few extra years of life by depriving yourself of today's pleasures ?" (Achilles problem) ; "to avoid health problems, do you pay

attention to your weight or your diet, do you practise a sport ?" ; or again, "do you prefer an increased public pension before age 85, in exchange of a reduced one thereafter ?".¹¹

Likewise, other questions reflect both impatience and time-preference, or both time preference and family altruism. As a general rule, we have decided to allow only for as much as *two* interpretation in terms of preferences, with time preference being always involved in such cases¹². Even with these limitations, the problem of multiple or ambiguous interpretations of questions in terms of revealed preferences has strong implications. The scores of risk tolerance, impatience, and altruism will share some questions in common with the score of time preference (even though different coding may be sometimes used), but these "polysemous" questions should not be set aside since they appear quite useful in the building of scores... This is one further reason to increase the number of questions, in order to limit the weight of overlapping questions.

(3) In order to give a priori equal weight to each question in the score, the same *coding*, using a rating scale of *three* response choice : -1 ; 0 ; +1, has been generally applied. For time preference, for instance, -1 corresponds to far-sighted, 0 to "neutral", and +1 to short-sighted ; for the risk score, -1 means more risk tolerant, 0 intermediate, and +1 to less risk-tolerant. Once again, there are several pitfalls. In psychometrics especially, it is typically the case that four to seven responses choices are offered to qualitative questions, as in the following five-level scale for opinions : strongly disagree, disagree somewhat, no position, agree somewhat, agree strongly (Spector, 1991). We checked on TNS-Sofres data that the results are robust to a change in the number of response choices offered, from three to five : the corresponding scores are highly correlated and, more importantly, there is almost no variations in the ranking of the individuals on each score.

Moreover, some answers are difficult to code for lack of further information on the respondent. How, notably, interpret the lack of practice of given risky activities ? The reasons may have nothing to do with the element of risk involved – coding 0 is then the better choice ; but the lack of practising may also reveal risk aversion – in this case, coding +1 is warranted. Without complementary information, there is bound to be errors of coding. The only hope is that by varying the context, the situations or the life domains involved, these errors would cancel for a large part.

(4) The last step concerns both the *calculation* and *internal validation* of scores for each individual. The "initial" value of each score is the sum, for the relevant questions, of the

¹¹ This is one further reason why our typology of savers, crossing scores of risk tolerance and time preference seems appropriate : the answers to the previous questions allow indeed to disentangle prudent and far-sighted "Conservative investors" from risk-tolerant and short-sighted "Hotheads" (see § 3.1).

¹² There is only one exception. The question : "In a couple where there is only one breadwinner, do you think it is important to cover financially the risk of his (or her) death, through life-insurance, appropriate savings... ?", appears simultaneously in the three scores of time preference, risk tolerance and altruism. Note that it would be the same with the choice between a lump-sum or an annuity payment (see above), which depends further of many more other uncontrollable factors.

"marks" obtained in the coding procedure. The basic assumption states that if most questions affected to risk (say) have a *common component* – presumably related to risk behaviour –, this component will be captured by the risk score, adequately adjusted. Statistical analysis will at the same time allow for a test of the existence of such a common component and give the adjusted value of the score. If the test is successful, the "final" value of the score will then be the same summation as before, but concerning only those items which are shown to form a coherent whole : questions with negative or too small correlation with the sum of all the others (the score without the question) will be eliminated, and Cronbach's coefficient alpha will be used to judge and maximise the degree of internal coherence of the measure (see Spector, 1991). The score is thus a summary measure, qualitative and ordinal, which is supposed to represent the individual's answers to a set of diverse questions.

We will be more explicit below (see § 5.3) on the procedure of aggregation that is more informative about the coherence of the indicator than about its substance : the risk score could as well capture, through the "common component", a personal traits that reflect something different from preference towards risk¹³. But the main lesson to be drawn so far is that steps 1 and 2 of the method – the design and affectation of questions to such or such preference – are the essential pieces for its success or the quality of scores, but that their evaluation can be made only *ex post*, while looking at the statistical properties of derived scores. In particular, the questions must be designed and grouped in such a way that their obvious shortcomings cancel through aggregation.

5.2. No one question is by itself satisfactory

Acknowledging that no single question, however carefully formulated, could bring reliable and accurate information of a given preference of the respondent is the hallmark of our empiricist and synthetic approach. It is true for theory-based abstract questions, and all the more so for the type of concrete, real-life questions we consider. It is useful to list the main objections that can be raised against their interpretation in terms of such or such preference, even when you accept step 2, i.e. the way each one has been a priori affected to risk tolerance, time preference, impatience or altruism.

(i) In many too anecdotic or specific situations, it is not possible to determine *which* preference parameter is precisely revealed by the question under consideration. Suppose thus that the respondent's reaction to the question "have you change your consumption patterns as a result of the Mad cow disease ?" reflects her behaviour towards risk or uncertainty. But what is measured ? Risk aversion, or prudence in the standard framework ? Or rather non-standard aversion to uncertainty, given that probabilities of contamination are largely unknown ? Or even irrational fears or panic ?

¹³ Precisely devoted to this issue, section 6 looks for miscellaneous "indirect" validations of scores as indicators of preference, comparing notably the scoring method with the results from principal components analysis.

(ii) There are few ways in many cases to get rid of *framing effects* and confounding factors : to the question "Do you often park in a forbidden zone ?", a risk-tolerant individual can still answer "No" owing to his / her feelings of good citizenship ;

(iii) A lot of questions show, alone, *little explanatory power of wealth* : the question "do you take an umbrella when the weather outlook is uncertain" has, hopefully, no correlation with savings, but has one of the highest contribution to the risk score (what is needed of a perfect instrument) ;

(iv) Finally, for the few questions that appear significantly correlated to wealth-related behaviours, the direction of causality is generally unclear, raising the issue of *endogeneity* : the question "do you have a hard time balancing your monthly budget ?", contributing to the scores of time preference and short-term impatience, does indeed predict (low) wealth, but there is obviously a problem of reverse causation.

Our key contention is that the method of aggregation through synthetic scores might, to a large extent, eliminate all these problems at the same time. Against (i) and (ii), the risk score (say) would thus be a context-free good summary of the scope of the individual's various preferences towards risk, *as if* it represented a "behavioural general disposition" towards risk – but only so on an ordinal basis. Against (iii), it should also have a significant explanatory power of wealth amount and composition, with econometric effects corresponding to theoretical predictions. And against (iv), the fact that it sums up the answers given to a lot of anecdotic or specific questions with little link with wealth should allow each score to pass the tests of exogeneity. Indeed, in similar tests regarding alternative measures of preferences, scores constitute instruments of higher quality and validity than other variables or specific questions designed for that purpose (e.g. planning vacations as an instrument of the propensity to plan in Ameriks *et al.*, 2003) : the main reason of that superiority is, once more, that they represent composite indicators based upon reactions to various situations in multiple life domains (see Appendix).

Put together, these propositions represent a very strong statement that can be called the "power of aggregation". Whether it is true or not is of course a matter of degree and essentially, at this point, an *empirical* issue. There is thus nothing like a central limit theorem that would allow us to treat framing effects or endogeneity biases as uncorrelated error terms that cancel by aggregation. But this analogy suggests at least that the statement has more chances to be true if the number of questions retained for each score is sufficiently high.

5.3. The scores : synthetic and ordinal measures of savers' preferences

Let us consider the different elements in the statement regarding the power of aggregation. Regarding (iii), we have already seen that scores have significant effects on wealth in the direction predicted by the theory (§ 4.3).

As for (iv), endogeneity issues are discussed at length in the Appendix. The results show that the preference scores can be considered as exogenous, so that the previous econometric effects are not significantly affected by causality bias. This conclusion is in a way not surprising, as the scores can be considered as the sum of a number of elements which can be considered as "natural" instruments (Angrist and Krueger, 2001). Regarding risk attitudes, the question about whether the individual "takes his/her umbrella if there is a chance of rain", which appears strongly correlated with the risk score, has no direct effect on the amount of wealth. Similarly, the "ability to forego current pleasure in order to live longer", which is strongly correlated with the time discount score, does not explain household assets¹⁴. Following on from the above, it is natural to think that the scores can be used as instruments for other preference parameters. The variables which are instrumented thus become, in a sense, "disguised" scores. It is thus shown in the Appendix that the risk-attitude and time-preference scores are very good instruments for the corresponding scales.¹⁵

What remained to be tested is whether the risk score (e.g.) captures a common component present in all – or most of – the questions affected to risk, and whether, moreover, this component represents a general disposition towards risk, a specific behavioural trait of the respondent.¹⁶

This brings us back to step 4 of our scoring method, the direct validation of scores and, more precisely, the test and evaluation of their degree of internal coherence (see § 5.1). Remember that the initial value of each score is the sum of the "points" obtained, through previous coding, for the questions affected to it.

The statistical approach followed is two-folds. First, we eliminate those questions that show a too small (or even negative) correlation with the sum of all other items – i.e. the initial score without that question. The correlation threshold was fixed at 5 %. If too many questions have to be rejected in this test, there is a flaw in our approach, whether in the choice and design of the questions or on their affectation (steps 1 and 2), or perhaps we should consider several heterogeneous scores for each preference. If only few or a limited number of questions have to be dropped, the score will then be the same summation concerning only those remaining items. Second, we try at the same time to evaluate and maximise the degree of

¹⁴ Only the family altruism score seems endogenous (at the six per cent level) in the financial wealth regression. This score is indeed constructed from a *smaller* number of elements, of which a number are directly related to wealth (notably transfers to children).

¹⁵ The instrumental variables results show that only the scale of time preference can be considered as endogenous (for financial, gross and net wealth). The OLS results for this scale are thus biased. Nonetheless, the IV results are in the same direction and are actually somewhat stronger : being far-sighted is positively correlated with wealth. We also considered IV estimation for scales *without* the scores as instruments. Instrument quality is sharply reduced, although it remains acceptable. But the exogeneity of the scales is no longer rejected in this specification, which illustrates the problems of estimation with weak instruments...

¹⁶ If there is such "behavioural specificity" (Barsky *et al.*, 1997), individuals will tend to be similarly ranked in their responses to many risky situations, covering different domains (family, health, social, work...), and one risk score will be sufficient, as a relative measure of risk preference, to capture this personal trait.

internal consistency of this purged score, using *Cronbach coefficient alpha*. Mainly used in psychometrics, this coefficient is equal to (see Spector, 1991) :

$$\kappa = \frac{n}{n-1} \left(1 - \frac{\sum_{i=1}^n \sigma_i^2}{\sigma^2} \right) \quad (7)$$

where n is the number of items (questions) introduced in the score, σ^2 represents the total variance of the score, and σ_i^2 the variance of item i .¹⁷

Note that a value of the Cronbach alpha coefficient below 0.40 is generally considered too low, at least in IQ tests – suggesting that a bunch of remaining questions may further eliminated, or that several "sub-scores" (e.g. per domain of life) should be considered. In such tests, internal consistency is judged fully satisfactory if the value of the coefficient exceeds 0.70 (see Nunnally, 1978).

Table 1 gives the results on Insee data for the four scores selected, one for each saver's preference to be measured. The results are quite satisfactory for the risk score : only two questions (out of 56) have been eliminated, and the resulting coefficient κ – on the remaining 54 questions – is high (0.65), granting especially the fact that preferences are highly subjective variables. They are still okay for time preference ($\kappa = 0.53$), allowing for the fact that this parameter raises a lot of conceptual difficulties (see Masson, 2000) ; indeed, the significant number of rejected questions is another sign of these difficulties. The results are poor for impatience – a large percentage of items are rejected (at the 5 % level of correlation), and coefficient κ is too low –, suggesting further elimination or sub-grouping of items ; but this cannot be done, owing to the limited number of questions concerning impatience in the questionnaire. Finally, we face a similar problem for family altruism, despite the fact that only one question is initially rejected : in this latter case, there is still hope that a larger number of questions or a more careful formulation of a number of them will bring better results¹⁸. Once again, a large number of questions is a basic requirement for a reliable aggregation procedure.

Take especially the case of risk. How to interpret that a single indicator is enough to capture the whole array of behaviours towards risk in various contexts or situations, in different life domains ? One reason is that we look only for rankings, not estimates in absolute terms. But more fundamentally, what does the common component detected by the score may well represent ? A unique underlying preference, a general behavioural disposition of each individual, which commands her homogeneous reactions in very heterogeneous settings ?

¹⁷ This coefficient κ equals zero for independent items. It is maximal, equal to 1, if all the questions are perfectly (positively) correlated. Its value rises, separately, with the number of items n and with the degree of covariance between the answers to the different questions. To maximise it, one has then to eliminate the items which are the least correlated with the others, but up to a certain point only (since κ decreases with n).

¹⁸ The score of altruism derived from TNS-Sofres data has indeed better statistical properties, owing to improved and more focused questions. On the other hand, results are less satisfactory for other scores due to a restricted number of questions.

Such a conclusion in favour of "behavioural specificity" appears a bit far-fetched and there are obvious counter-examples. The (French) world-recorder of dragster – a very risky sport – admitted recently that he had otherwise a very wise and prudent life, concerning health, financial investments and the like : such foresight and prudence were in fact a necessary condition for him to continue his risky professional activity. Yet, data say that such a case has limited weight on a statistical basis. Indeed, Table 2, giving the correlations between risk *sub-scores* per domain on life, shows always positive, if limited, correlations between sub-scores (often below .2) and, more importantly, high positive correlations (above .5) between the global risk score and each sub-score. These results tell a similar story as the high value of internal consistency of the risk score, as measured by Cronbach coefficient alpha.

A principal component analysis offers still another and complementary way to check how a single risk score can represent a reliable summary of the answers given to more than fifty heterogeneous questions. The active variables are the initial 56 coded questions. Graph 3 represents their projections in the plan of the two first axes which explain only 11.2 % (6.1 % + 5.1 %) of total dispersion – but still much more than the following axes. The score is projected in quartiles on the same plan, with SAR1 (in the North-East) for the more prudent, and SAR4 (in the South-East) for the most risk-tolerant. This projection, a largely spread out and almost straight line which goes approximately through the origin, shows that the score is a balanced and extended combination of information brought by the two first axes ; moreover, it is quasi-orthogonal to the third axis (explaining only 3.6 % of total dispersion). By comparison, the projection of the risk *scale* – respondent's self-perception between 0 (prudent) and 10 (adventurous) –, also in quartiles, is much nearer to the origin and conveys information of the first axis only ; on the other hand, the two projections are, hopefully, positively correlated. These results illustrate the large superiority of scores over self-declared indicators of preference.

6 – INDIRECT VALIDATIONS OF PREFERENCE SCORES

Granted that scores have nice statistical properties, capturing a common component of many different questions, the last issue is about substance. Do they represent a reliable indicator of the corresponding preference, or something else ? The scattered pieces of evidence given below do indeed converge to a strongly favourable assessment ; yet, they constitute nothing more than "indirect" confirmations. In the absence of a structural model, no direct tests of the validity of scores can be envisaged (see the earlier discussion in § 4.2.).

Consider again the principal component analysis concerning risk preference (Graph 3). We have just seen that the risk score is bi-dimensional, capturing a rich mixture of the information conveyed by the two first axes, but none of the third one. But what interpretation

can be given to each axis ? The first one corresponds to everyday life choices ("does not park in a forbidden zone", "arrives well in advance when taking a train or a plane"...) and those relating to early economic life, marital or professional ("marriage is an insurance", "keeps a close eye on young or teenaged children, teach them to limit risks taken", "prefers a safe job"...). The second axis concentrates rather on longer-term and life risks, especially concerning health and retirement ("worries about ending up her life in a nursing home, and has put some money aside to face this eventuality", "feels concerned by the debates on the financing of public health expenditures", "worries about being in good shape – exercises, weight and diet watching", "in a couple when there is only one breadwinner, thinks it is important to protect the partner in case of death"...). On the other hand, the meaning of the third axis appears unclear. These interpretations show once more the richness of the risk score, and the bias of self-declared scales, which seem to concentrate on only one type of risks (current or early ones).

Comparable qualitative conclusions can be drawn for the other scores. The latter appear to sum up nicely the common information brought by the various questions. Moreover, they appear much more informative than the corresponding scales but, hopefully, the two types of indicators point in related directions.¹⁹

Correlations between scores and corresponding scales confirm this last point : as indicated on Table 3 on Insee data, they are always significant and largely positive, albeit far inferior to one – scales remain poor proxies for scores. Note that there is no scale for altruism, as we did not know how interviewees could really position themselves on such a scale, between 0 (totally selfish) and 10 (altruistic to the last degree, like Balzac's Père Goriot ?).

Another field where there is large agreement between our measures of preferences and self-perception or beliefs about one's (or others') preferences concerns age and gender effects on scores. Consider first age effects. We find on a cross-section that, according to scores, prudence and foresight increase with age (effects are similar, but less important, for other measures, such as scales). Interviewees were asked whether they had the impression that their attitude towards risk or time preference changed substantively during the course of their life. Regarding risk, 48 % answered their attitude towards risk changed, with more than 42 % saying they become more prudent, and less than 6 %, only, less prudent. Regarding time preference, 44 % answered that changed foresight, with almost 40 % saying they become more far-sighted, and only 4.2 % less farsighted. Thus, in both cases, an overwhelming majority of the respondents who consider themselves to have changed preferences say they

¹⁹ For instance, the time preference score is also bi-dimensional, capturing the information of the two first axes : its projection on the plan is again almost a straight line which goes approximately through the origin and is close to the bisecting line ; the projected time preference scale is much more irregular and nearer to the origin, although it points, on the average, to the same direction. The first axis corresponds to a weak sensitivity to long-term social problems (sustainability of the retirement system, environment), and the second, to persons unconcerned or carefree about their own future (financial budget, standard of living in retirement)... or that of their children (education, taste for saving).

that they have become more prudent and/or far-sighted, so that cross-sectional age effects on scores should be first interpreted as life-cycle effects.²⁰

Consider then gender effects. According to scores, men are less prudent than women (a general result in the literature) but also no less far-sighted (gender does not explain the time preference score, neither the time preference scale). For couples, a question at the end of the questionnaire asks where do you place your spouse regarding risk attitude or time preference. In this self-assessment of spouse's preferences, both husbands and wives agree, in large proportions, that the wife is more prudent. But both tend also to consider their partner to be the more far-sighted, a "contradiction" suggesting that the sex effect on time preference is indeed ambiguous.²¹

Another indirect way to evaluate the quality of the scores relatively to other measures of preferences is to compare their explanatory power of risky activities – or others involving long-term foresight – while taking off these activities from the corresponding score. We thus check that the risk score, adequately "purged", explains playing games (P.M.U., Loto, Slot machine, casino), or self-declared preference for risky investment, better than risk scales or Barsky *et al.* measure of risk tolerance (Arrondel and Masson, 2007, chapter 6).

7 – CONCLUSIONS

How much wealth inequality or households' portfolio diversity can be attributed to heterogeneous preferences of savers towards risk or time ? The answer to this question involves a tricky part : the evaluation on individual preferences based upon microeconomic theory of saving, i.e. some form of an "extended" life-cycle model. Here, most previous attempts, generally based on one or two abstract questions, have been resounding failures, especially regarding our main goal, that is assessing the explanatory power of savers' preferences on wealth-related behaviours.

For that reason, we tried a completely different method of measurement : we multiply the number of more concrete, real-life questions of various types and covering a large array of life domains, while forgetting the hope to get rid, *ex ante*, of framing effects and confounding factors : these polluting elements should be eliminated through aggregation in synthetic and ordinal scores. This piecemeal and empiricist approach has proved successful at different levels : parsimony is saved, since data tell, for instance, that one risk score is enough for our purpose ; scores of preferences do significantly explain wealth in the direction predicted by

²⁰ It is not possible to draw firmer conclusions without panel data, that will be available over five years for more than a thousand individuals in our next survey, now running (TNS-Sofres 2007).

²¹ More surprising at first sight is our result that (French) men are no less altruistic towards children than women, and this holds even in the sample of couples with children. Here, no self-declared scale or self-assessment of the spouse's preference is available. We check that the result cannot be imputed to the aggregation procedure. In the new survey (TNS-Sofres 2007), we will have independent measures of scores, including altruism, for both spouses.

theory ; and we have indirect confirmation that scores are reliable indicators of the preferences that ought to measure, and not of something else.

Now, our method remains full of frailties and must be replicated on different data sets to check its robustness. In that perspective, a key lesson drawn when comparing Insee 1998 (and 2004) and TNS-Sofres 2002 surveys is that a sufficiently *large* number of heterogeneous questions covering many domains is required. There is no short cut : whenever you try to keep only a few questions, while selecting the "best ones" or while concentrating in one domain of life (e.g. financial or professional), scores gradually lose their good properties : internal consistency, explanatory power of wealth-related behaviours, exogeneity in wealth regressions, and especially robustness to alternate choices of questions.²²

Our future agenda follows the same line of research, while trying to solve a number of problems raised in this paper. The new survey, TNS-Sofres 2007, will thus include a panel (from 2002 to 2007) in order, among other things, to test if scores represent durable preferences and not too volatile perceptions or beliefs, regarding risk especially. There are more questions devoted to altruism and social issues. The larger size of the survey will also help us to refine our typology of savers by allowing for more crossing of preferences.

Moreover, independent measures of scores for each spouse should be available in a number of couples, allowing to identify mating patterns in terms of preferences towards risk and time. We know that there is strong assortative mating in terms of income, social class, etc., but these variables have only a limited explanatory power on (scores of) preferences, so that we know little, still, on how people match in terms of risk attitude, time preference, altruism or short-term impatience. This is an important issue since we have, up to now, related household's wealth to the scores of *one* spouse only. We will thus have, as an intermediate step, to consider the following question, typically addressed by models of household collective decision making : what is the outcome of saving and portfolio intra-household decisions when the preferences of the two spouses differ ?

Finally, we plan to compare our estimates of scores with the results obtained, for the same individuals, by standard experimental work (see Cohen *et al.*, 2007).

²² For instance, results obtained in the two surveys are very close if the same set of questions is used in each case : for instance, the two diagrams of principal component analysis are remarkably similar. When the two sets of questions do not coincide, differences become significant if the number of (common) questions is low, but remain minor if a large number of questions is used (as in the risk score).

Bibliography

Ameriks J., Caplin A. and Leahy J. (2003), "Wealth Accumulation and the Propensity to Plan", *Quarterly Economic Journal*, 118 (3), p. 1007-1047.

Angrist J. and Krueger A. (2001), "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments", *Journal of Economic Perspectives*, 15 (4), p. 69-87.

Arrondel L. (2002), "Risk Management and Wealth Accumulation Behavior in France", *Economics Letters*, n° 74, p. 187-194.

Arrondel L. and Calvo-Pardo H. (2002), "Portfolio Choice with a Correlated Background Risk: Theory and Evidence ", Paris, Cahier Delta n° 02-16.

Arrondel L. and Masson A. (2006), "Altruism, Exchange or Indirect Reciprocity: What Do the Data on Family Transfers Show?", à paraître in *Handbook on the Economics of Giving, Reciprocity and Altruism*, L. Gerard-Varet, J. Mercier-Ythier J. and S. C. Kolm (éds), North-Holland, Amsterdam, Chapitre 14, p. 971-1053.

Arrondel L. and Masson A. (2007), *Inégalités patrimoniales et choix individuels : des goûts et des richesses*, Economica, Paris, 397 pages.

Arrondel L., Masson A. and Verger D. (2004), "Préférences de l'épargnant et accumulation patrimonial", (dossier de 5 articles), *Économie et Statistique*, n° 374-375.

Arrondel L., Masson A. and Verger D. (2005), "Préférences face au risque et à l'avenir : types d'épargnants", *Revue Économique*, n°56 (2), p.393-416.

Arrow K.J. (1965), *Aspect of the Theory of Risk Bearing*, Yrjö Johnson Lectures, Helsinki.

Barsky R.B., Kimball M.S., Juster F.T. and Shapiro, M.D. (1997), "Preference Parameters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Survey", *Quarterly Journal of Economics*, 112 (2), p. 537-580.

Becker G.S. and Mulligan C.S. (1997), "On the Endogenous Determination of Time Preference", *Quarterly Economic Journal*, 112 (3), p. 729-758.

Bound J., D. Jaeger A. and Baker R.M. (1995), "Problems with Instrumental Variables Estimation when the Correlation Between the Instruments and the Endogenous Explanatory Variable Is Weak", *Journal of the American Statistical Association*, 90 (430), p. 443-450.

Caroll C.D. (2001), "A Theory of the Consumption Function, with and without Liquidity Constraints", *Journal of Economic Perspectives*, 15 (3), p. 23-45.

Cohen M., Tallon J.M. and Vergnaud J.C. (2007), "Choice under imprecise information: Experimental evidence", mimeo.

Fisher I. (1930), *The Theory of Interest*, Macmillan, New York.

Frederick S., Loewenstein G and O'Donoghue T. (2002), "Time Discounting and Time Preference: a Critical Review", *Journal of Economic Literature*, 40, p. 351-401.

Kahneman D and Tversky A (1979), "Prospect Theory: An Analysis of Decision under Risk", *Econometrica*, 47, p. 263-291.

Kapteyn A. and Teppa F. (2002), "Subjective Measures of Risk Aversion and Portfolio Choice", CENter Discussion paper n°2002-11, Tilburg.

Kreps D.M. and Porteus E.L. (1978), "Temporal Resolution of Uncertainty and Dynamic Choice Theory", *Econometrica*, 46 (1), p. 185-200.

Laibson D. (1997), "Golden Eggs and Hyperbolic Discounting", *Quarterly Journal of Economics*, 112, p. 443-477.

Lusardi A. (2003), "Planning and Savings for Retirement", Working Paper, Dartmouth College.

Masson A. (2000), "L'actualisation du futur", *Le genre humain*, 35, (n° spécial : *Actualités du contemporain*), p. 197-244.

Modigliani F. (1986), "Life Cycle, Individual Thrift and the Wealth of Nations", *American Economic Review*, 76 (3), p. 297-313.

Nunnally J.C. (1978), *Psychometrics Theory*, 2nd ed., McGraw-Hill, New York.

Robin J.-M. (1999), "Modèles structurels et variables explicatives endogènes", *Méthodologie statistique*, n° 2002, INSEE, Paris.

Spector P.E. (1991), *Summated Rating Scale Construction: an Introduction*, Sage.

Starmer C. (2000), "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk", *Journal of Economic Literature*, 38 (2), p. 332-382.

Venti S.F. and Wise D.A. (2001), "Choice, Chance and Wealth Dispersion at Retirement", in S. Ogura, T. Tachibanaki, and D. Wise (eds), *Aging Issues in the United States and Japan*, University of Chicago Press.

Appendix

Individual Preferences and Wealth Accumulation: the Problem of Causality

One of the major contribution of our study concerns the effect of individual preferences on wealth (Arrondel *et al.*, 2004). But in this analysis, especially in wealth equations, we cannot ignore the question of reverse causality between preferences and wealth.

Standard models of saving and portfolio choice often include the hypothesis of absolute risk aversion decreasing in wealth (Arrow, 1965). Recent theories have also linked lower time depreciation to higher wealth (Becker and Mulligan, 1997). Reverse causality is therefore a potential problem in equation (2) : the rich may take more risks, have more foresight, or again be more altruistic with respect to their children. Simple regression results are therefore difficult to interpret : the negative effect of time preference and the positive effect of family altruism, especially, may be statistical artefacts rather than causal relationships.

We therefore re-estimate our wealth equations using instrumental variables to evaluate the robustness of our results, and to test for the endogeneity of our preference scores.

Instrumental variables and endogeneity tests

The wealth equation is:

$$W_i = X_{1i} b_1 + X_{2i} b_2 + u_i = X_i b + u_i \quad (\text{A})$$

where X_1 denotes the vector of potentially endogenous preference parameters, X_2 the exogenous variables and u a normally- distributed error term.

The usual solution for causality problems is instrumental variables (IV) estimation, where the instruments are supposed to be orthogonal to the dependent variable, W .

The method of Two-Stage Least Squares (2SLS) provides a class of IV estimators in two stages. First, one or more instruments (denoted hereafter by the vector Z_1) provide, with the other exogenous variables X_2 , an OLS "prediction" of the potentially endogenous explanatory variables :

$$X_{1i} = Z_{1i} \alpha_1 + X_{2i} \alpha_2 + v_i = Z_i \alpha + v_i. \quad (\text{B})$$

Second, the predicted values (\hat{X}_1) are substituted into equation (A) which is re-estimated by OLS:

$$W_i = \hat{X}_{1i} b_{12MC} + X_{2i} b_{22MC} + w_i. \quad (\text{C})$$

If the instruments are correlated with the potentially endogenous explanatory variables, but orthogonal to the dependent variable, the 2SLS estimator is consistent. In practice, we have to check (see Tables E) both the quality (correlation between X_1 and Z , and between X_1 and Z_1 in equation (B)) and the statistical validity of the instruments

(orthogonality between Z_1 and u) from the following regression, where the coefficients μ should not be significantly different from zero:

$$\hat{w}_i = Z_{1i}\mu + \varepsilon_i. \quad (\text{D})$$

The 2SLS estimators also allow to test the endogeneity of the explanatory variables concerned (Robin, 1999). The usual test of Hausman (1978) consists in comparing the two estimators (OLS and 2SLS). This test is equivalent to carrying out, by OLS, an "augmented" regression in which the dependent variable, W , is estimated as a function of all the explanatory variables (X) and of the residuals (\hat{v}) estimated in the instrumental equation (B) from the first stage μ (see Holly, 1983) :

$$W_i = X_i b + \hat{v}_i d + u_i. \quad (\text{E})$$

If the variables in question are exogenous, the estimated coefficients d on the residuals will be insignificant (we can show in fact that $d = \text{Cov}(u, v) / \text{Var } v$) ; if the estimated coefficients are significant, then exogeneity is rejected (see Tables E).

The instrumentation of individual preferences : scores, scales and "plans"

Many of the preference measures are potentially endogenous in our wealth equations : the scores themselves, the scales, or even the single questions which are supposed to measure certain taste parameters (such as the propensity to establish long-term projects).

The main difficulty is finding good instruments for preferences which are independent of wealth. When instruments are weak, "the cure can be worse than the disease" (Bound *et al.*, 1995).

a) Scores

Table E1 summarises the results from endogeneity tests of the scores of risk and time preferences and family altruism.

The instruments used reflect characteristics of the *parents* of the respondent (social class, wealth composition, money problems, preferences over risk and time) and the existence of gifts given by the household¹. In all of the regressions, the instruments are significantly correlated with the scores and largely pass the validity (or over-identification) tests. While the R^2 statistic (of the instruments) is not particularly high (between 4 % and 12 %), it is nonetheless much larger than that resulting from endogeneity tests in American studies (Lusardi, 2003, and Ameriks *et al.*, 2003).

All of the scores bar one pass the exogeneity test. The residuals from the instrumental equations are never significant, except that for family altruism which is significant at the 6 % level in the financial wealth equation. We then conclude that the OLS estimates of equation (A) do not suffer from causality bias.

¹ Using parental preference variables as instruments is analogous to the practice in panel econometrics of using lagged variables (Robin, 1999).

b) The scales of risk-aversion and time preference

Two series of regressions are analysed here. The first is analogous to that above, applied to the two scales (of risk-aversion and time preference). The second uses, in addition, the scores as natural instruments to explain the scales.

The results of the first series of regressions are shown in Table E2. The conclusions are similar to those from the analysis of the scores. For the instruments which pass the quality and validity tests, the two scales seem exogenous ; the OLS estimators are thus unbiased.

The conclusion changes sharply when we instrument the scales with the scores (Table E3). We first note that the introduction of the scores considerably improves the quality of the instrumentation : the partial R^2 coefficients are now between 14 % and 21 % with F-statistics multiplied roughly by a factor of five. Over-identification tests show that the instruments are orthogonal to the residual of equation (A). However, contrary to the results in the first series of regressions, we now reject, for all three measures of wealth, the hypothesis of exogeneity in the case of time preference.

The IV or 2SLS estimators yield a positive correlation between the three wealth measures and the fact of considering oneself as preoccupied by the future on a scale of 0 to 10. The quantitative effect of this variable is however almost four times larger in 2SLS than OLS.

c) Plans (from 0 to 30 years)

A number of American studies (Lusardi, 2003, and Ameriks *et al.*, 2003) have looked at the relationship between an ability to plan for the future (rather than an indicator of time preference) and wealth. They use one or two subjective questions, which are supposed to measure this propensity. Ameriks *et al.* (2003) specifically consider long-term financial planning for retirement ; Lusardi (2003) looks at worries expressed about retirement.²

In a similar way, we can use a question in our survey asking individuals if they have (or had) made plans in a number of different areas of their life (career, family leisure time, wealth etc.) over a period of 10, 20, or 30 years or more. In the OLS equations for gross and net wealth, this question attracts a statistically significant coefficient (Table E4).

To instrument planning³, we use in the first instance only the time preference of the respondent's mother (see Table E4). This instrument passes the tests of quality and validity,

² In the two cited articles, the authors consider that the variable is endogenous, so that they have to turn to instrumentation. They do not carry out exogeneity tests. Ameriks *et al.* (2003) instrument the propensity to plan by other subjective questions supposed to be less correlated with wealth. These concern preparation for holidays, and confidence in one's ability to calculate. Lusardi (2003) instruments "worrying about retirement" by the difference in age between the respondent and his/her eldest brother or sister, and by parents' health, the idea being that one benefits from one's siblings' or parents' experience.

³ Since the variable to be instrumented is ordinal (the planning horizon in four bands) we should use an ordered Probit for the instrumental regression. Angrist and Krueger (2001) show, however, that a qualitative model does not produce consistent estimators if the specification is inexact.

but even so the correlation with the planning variable is weak (0.6 %). We therefore have weak instruments which can produce biased IV estimators... as in the two American studies cited (a correlation between 0.2 and 1.2 % in Lusardi, 2003, for example)⁴. Exogeneity is only rejected for gross wealth. The IV estimator, positive, is much larger than its OLS counterpart.

In Table E5 we use the time preference score as an instrument for planning. Instrument quality rises sharply (the partial R^2 is multiplied by 10 and the F -statistic by 7). Exogeneity is rejected in all of the wealth regressions. We still find the positive relation between the time-length of planning and household wealth in the IV results, with a quantitatively larger correlation than in the OLS estimations. Note that, for gross wealth, the difference between the IV and OLS estimators is smaller when we use the score as an instrument. The difference between the two IV estimators can be traced back to the weakness of the instruments other than the score (Bound *et al.*, 1995, p. 444).

⁴ Ameriks *et al.* (2003) only show the F -statistic of the instruments, which is very significant (<0.0001). However, orthogonality (via the over-identification test) is only accepted at the 8% level.

Table E1. The Instrumentation of Scores

<i>Variables</i>	<i>Log Financial Wealth</i>		<i>Log Gross Wealth</i>		<i>Log Net Wealth</i>	
	<i>Statistic</i>	<i>Critical Level</i>	<i>Statistic</i>	<i>Critical Level</i>	<i>Statistic</i>	<i>Critical Level</i>
<i>Exogeneity Tests in Augmented Regressions (t-statistic)</i>						
Risk-Aversion	-0,64	0,521	0,56	0,578	0,22	0,829
Time Preference	-0,33	0,740	0,53	0,600	1,15	0,253
Family Altruism	-1,87	0,062	-0,95	0,341	0,33	0,739
<i>Instrument Quality</i>						
Risk-Aversion						
R ²	0,297		0,289		0,301	
F	6,380	< 0,0001	7,01	< 0,0001	7,23	< 0,0001
R ² (instruments only)	0,115		0,072		0,053	
F(instruments only)	7,18	< 0,0001	7,19	< 0,0001	6,55	< 0,0001
Time Preference						
R ²	0,170		0,162		0,161	
F	3,09	< 0,0001	3,34	< 0,0001	3,23	< 0,0001
R ² (instruments only)	0,066		0,061		0,056	
F(instruments only)	3,93	< 0,0001	6,05	< 0,0001	6,90	< 0,0001
Family Altruism						
R ²	0,183		0,176		0,179	
F	3,40	< 0,0001	3,68	< 0,0001	3,68	< 0,0001
R ² (instruments only)	0,075		0,044		0,038	
F(instruments only)	4,46	< 0,0001	4,29	< 0,0001	4,60	< 0,0001
<i>Instrument Validity</i>						
F	0,64	0,882	1,25	0,246	1,13	0,338
<i>Number of Observations</i>						
	1129		1131		1051	

Source: INSEE Patrimoine 1997 Survey

Instruments. For financial wealth: parents' preferences, money problems when young, parents' social class, bequests or gifts given (20 variables). For gross wealth: parents' preferences, money problems when young, parents' social class, bequests or gifts given (12 variables). For net wealth: parents' preferences, bequests or gifts given (9 variables).

Table E2. The Instrumentation of Scales

<i>Variables</i>	<i>Log Financial Wealth</i>		<i>Log Gross Wealth</i>		<i>Log Net Wealth</i>	
	<i>Statistic</i>	<i>Critical Level</i>	<i>Statistic</i>	<i>Critical Level</i>	<i>Statistic</i>	<i>Critical Level</i>
<i>Exogeneity Tests in Augmented Regressions (t-statistic)</i>						
Risk-Aversion	1,34	0,18	-0,75	0,452	-0,33	0,739
Time Preference	0,40	0,69	-1,27	0,206	-0,80	0,424
<i>Instrument Quality</i>						
Risk-Aversion						
R ²	0,164		0,148		0,161	
F	2,61	< 0,0001	2,62	< 0,0001	2,67	< 0,0001
R ² (instruments only)	0,080		0,058		0,065	
F(instruments only)	4,37	< 0,0001	5,26	< 0,0001	5,44	< 0,0001
Time Preference						
R ²	0,130		0,119		0,119	
F	1,99	< 0,0001	2,03	< 0,0001	1,89	< 0,0001
R ² (instruments only)	0,052		0,041		0,044	
F(instruments only)	2,75	< 0,0001	3,62	< 0,0001	3,62	< 0,0001
<i>Instrument Validity</i>						
F	0,66	0,863	0,77	0,687	0,76	0,697
<i>Number of Observations</i>						
	1030		1031		959	

Source: INSEE Patrimoine 1997 Survey

Instruments. For financial wealth: parents' preferences, money problems when young, parents' social class, bequests or gifts given (20 variables). For gross wealth: parents' preferences, money problems when young, parents' social class, bequests or gifts given (12 variables). For net wealth: parents' preferences, bequests or gifts given (9 variables).

Table E3. Instrumentation of scales including scores

<i>Variables</i>	<i>Log Financial Wealth</i>		<i>Log Gross Wealth</i>		<i>Log Net Wealth</i>	
	<i>Statistic</i>	<i>Critical Level</i>	<i>Statistic</i>	<i>Critical Level</i>	<i>Statistic</i>	<i>Critical Level</i>
<i>Exogeneity Tests in Augmented Regressions (t-statistic)</i>						
Risk-Aversion	0,54	0,59	-0,59	0,556	-0,15	0,878
Time Preference	-2,23	0,03	-3,38	< 0,001	-2,37	0,018
<i>Instrument Quality</i>						
Risk-Aversion						
R ²	0,255		0,256		0,258	
F	5,35	< 0,0001	5,36	< 0,0001	5,04	< 0,0001
R ² (instruments only)	0,206		0,206		0,204	
F(instruments only)	26,40	< 0,0001	26,46	< 0,0001	24,23	< 0,0001
Time Preference						
R ²	0,187		0,187		0,195	
F	3,60	< 0,0001	3,59	< 0,0001	3,50	< 0,0001
R ² (instruments only)	0,137		0,137		0,148	
F(instruments only)	16,22	< 0,0001	16,24	< 0,0001	16,41	< 0,0001
<i>Instrument Validity</i>						
F	0,59	0,827	0,55	0,855	0,41	0,944
<i>Estimators of the scale of time preference</i>						
OLS	0,061 (0,019)	< 0,001	0,081 (0,021)	< 0,001	0,068 (0,021)	< 0,001
IV	0,222 (0,082)	< 0,001	0,349 (0,092)	< 0,001	0,247 (0,085)	< 0,001
<i>Number of Observations</i>						
	1029		1031		959	

Source: INSEE Patrimoine 1997 Survey

Instruments used: Time Preference and Risk-Aversion Scores, Parents' preferences (10 variables).

Table E4. Instrumentation of plans (from 0 to 30 years)

<i>Variables</i>	<i>Log Financial Wealth</i>		<i>Log Gross Wealth</i>		<i>Log Net Wealth</i>	
	<i>Statistic</i>	<i>Critical Level</i>	<i>Statistic</i>	<i>Critical Level</i>	<i>Statistic</i>	<i>Critical Level</i>
<i>Exogeneity Tests in Augmented Regressions (t-statistic)</i>	-1,64	0,101	-2,39	0,017	-1,560	0,119
<i>Instrument Quality</i>						
<i>R</i> ²	0,102		0,100		0,099	
<i>F</i>	2,22	< 0,0001	2,21	< 0,0001	2,04	< 0,0001
<i>R</i> ² (instruments only)	0,007		0,006		0,006	
<i>F</i> (instruments only)	3,77	< 0,02	3,51	< 0,03	3,32	< 0,04
<i>Instrument Validity</i>						
<i>F</i>	0,21	0,812	0,11	0,895	0,13	0,876
<i>Estimators</i>						
<i>OLS</i>	0,008 (0,005)	0,14	0,012 (0,006)	0,04	0,013 (0,006)	0,03
<i>IV</i>			0,213 (0,122)	0,08		
<i>Number of observations</i>	1030		1131		1052	

Source: INSEE Patrimoine 1997 Survey

Note: The dependent variable takes four values (no plans, plans over 10, 20, or 30 years or more)

Instruments used: Mother's Time Preference (2 variables).

Table E5. Instrumentation of plans (from 0 to 30 years) including the score

<i>Variables</i>	<i>Log Financial Wealth</i>		<i>Log Gross Wealth</i>		<i>Log Net Wealth</i>	
	<i>Statistic</i>	<i>Critical Level</i>	<i>Statistic</i>	<i>Critical Level</i>	<i>Statistic</i>	<i>Critical Level</i>
<i>Exogeneity Tests in Augmented Regressions (t-statistic)</i>	4,680	< 0,0001	-4,180	< 0,0001	-3,570	< 0,001
<i>Instrument Quality</i>						
<i>R</i> ²	0,167		0,166		0,171	
<i>F</i>	3,85	< 0,0001	3,89	< 0,0001	3,74	< 0,0001
<i>R</i> ² (instruments only)	0,066		0,066		0,075	
<i>F</i> (instruments only)	26,49	< 0,0001	26,35	< 0,0001	28,43	< 0,0001
<i>Instrument Validity</i>						
<i>F</i>	0,24	0,869	0,76	0,517	0,31	0,821
<i>Estimators</i>						
<i>OLS</i>	0,008 (0,005)	0,14	0,012 (0,006)	0,04	0,013 (0,006)	0,03
<i>IV</i>	0,090 (0,021)	< 0,0001	0,096 (0,023)	< 0,0001	0,081 (0,021)	< 0,001
<i>Nombre d'observations</i>	1030		1131		1052	

Source: INSEE Patrimoine 1997 Survey

Note: The dependent variable takes four values (no plans, plans over 10, 20, or 30 years or more)

Instruments used: Mother's Time Preference and time preference score (3 variables).

Table 1 : Scores properties

Scores	Cronbach's Alpha			Final Items/Initial Items
	<i>Total population</i>	<i><= 40 years old</i>	<i>> 40 years old</i>	
<i>Risk</i>	0,65	0,62	0,62	54/56
<i>Time Preference</i>	0,53	0,44	0,56	25/34
<i>Impatience</i>	0,27	0,22	0,32	8/13
<i>Familial Altruism</i>	0,29	0,23	0,33	8/9

Source: Patrimoine 1998, Insee-Delta survey.

Table 2 : Correlations between risk subscores

	<i>Global score</i>	<i>Consumption</i>	<i>Labour</i>	<i>Financial managing</i>	<i>Health</i>	<i>Family</i>	<i>Retirement</i>	<i>Others</i>
<i>Global score</i>	1,00	0,56	0,48	0,44	0,56	0,68	0,49	0,56
<i>Consumption</i>		1,00	0,22	0,10	0,12	0,19	0,11	0,21
<i>Labour</i>			1,00	0,03	0,08	0,20	0,14	0,22
<i>Financial managing</i>				1,00	0,21	0,23	0,23	0,11
<i>Health</i>					1,00	0,22	0,25	0,21
<i>Family</i>						1,00	0,29	0,18
<i>Retirement</i>							1,00	0,23
<i>Others</i>								1,00

Source: Patrimoine 1998, Insee-Delta survey.

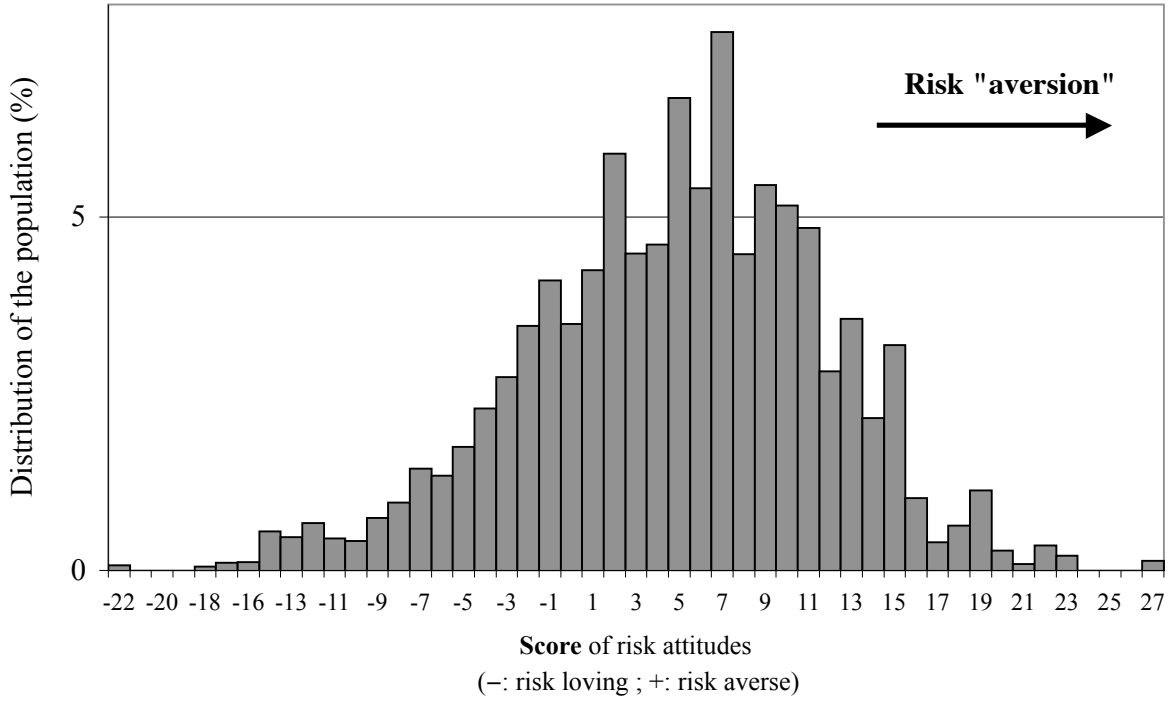
Table 3 : Correlation between scores and scales

	<i>Correlations</i>
<i>Risk</i>	0,44
<i>Time Preference</i>	0,28
<i>Impatience</i>	0,33

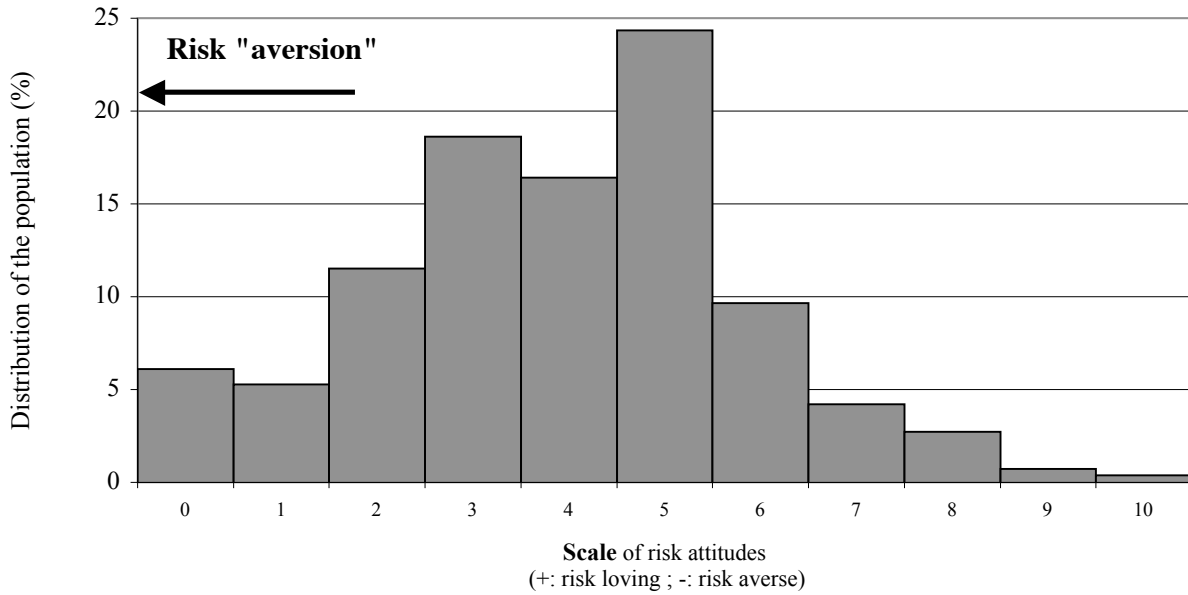
Source: Patrimoine 1998, Insee-Delta survey.

Graph 1 : Distribution of the population according to risk attitudes

a. Score

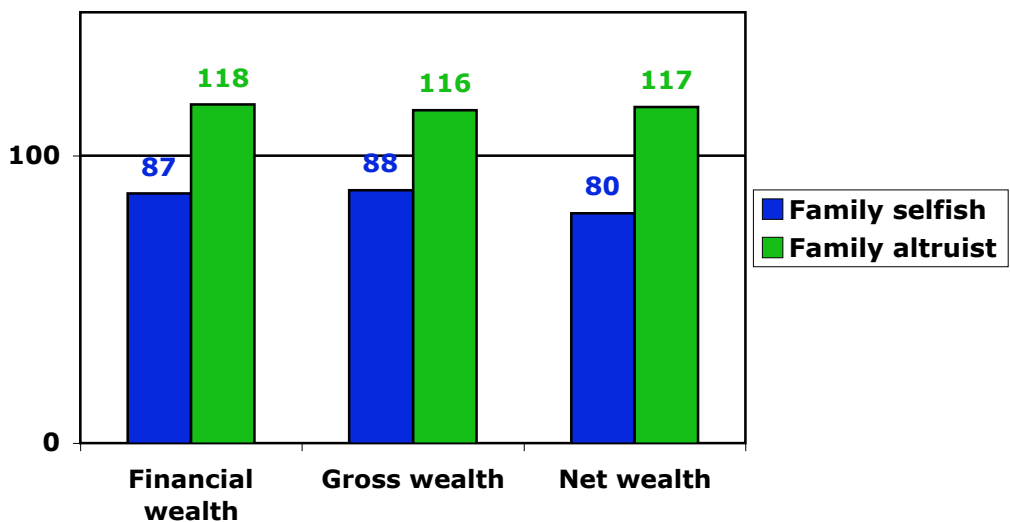
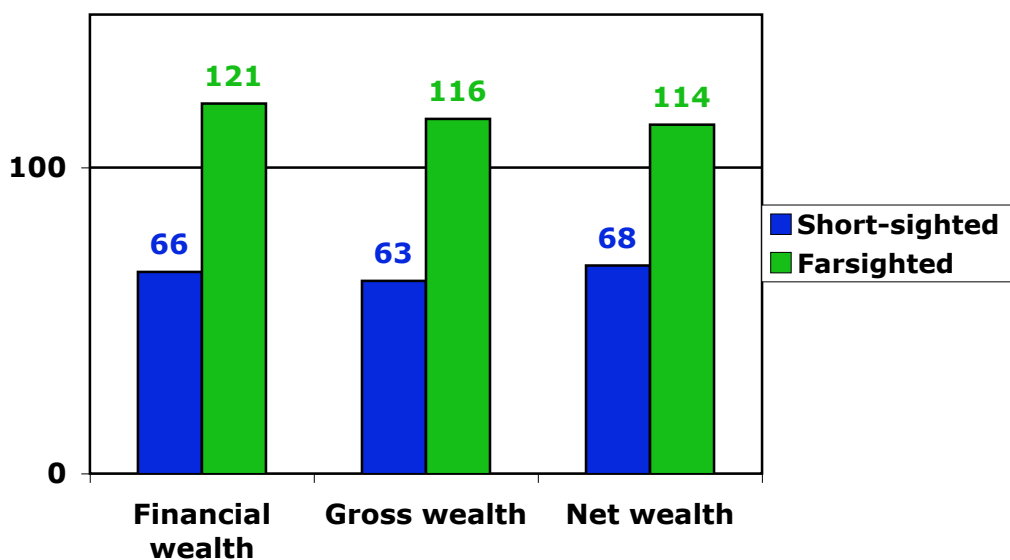
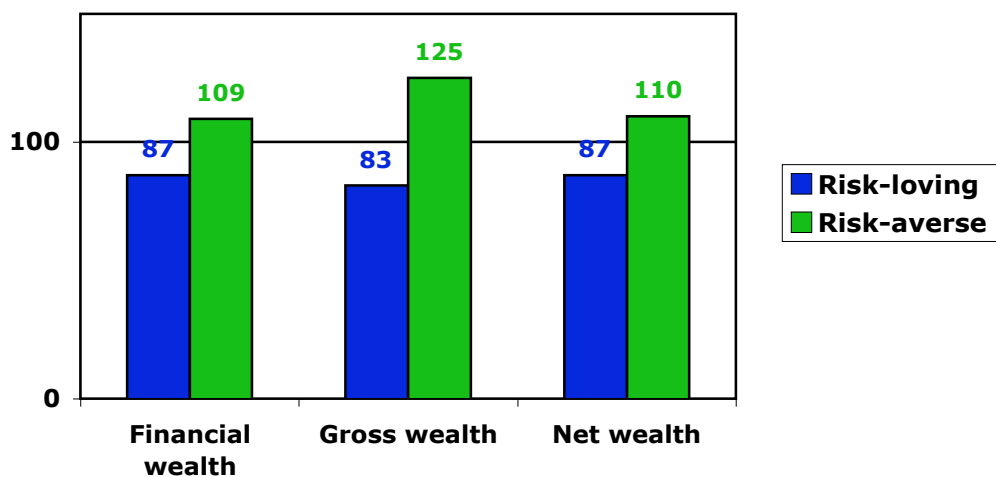


b. Scale



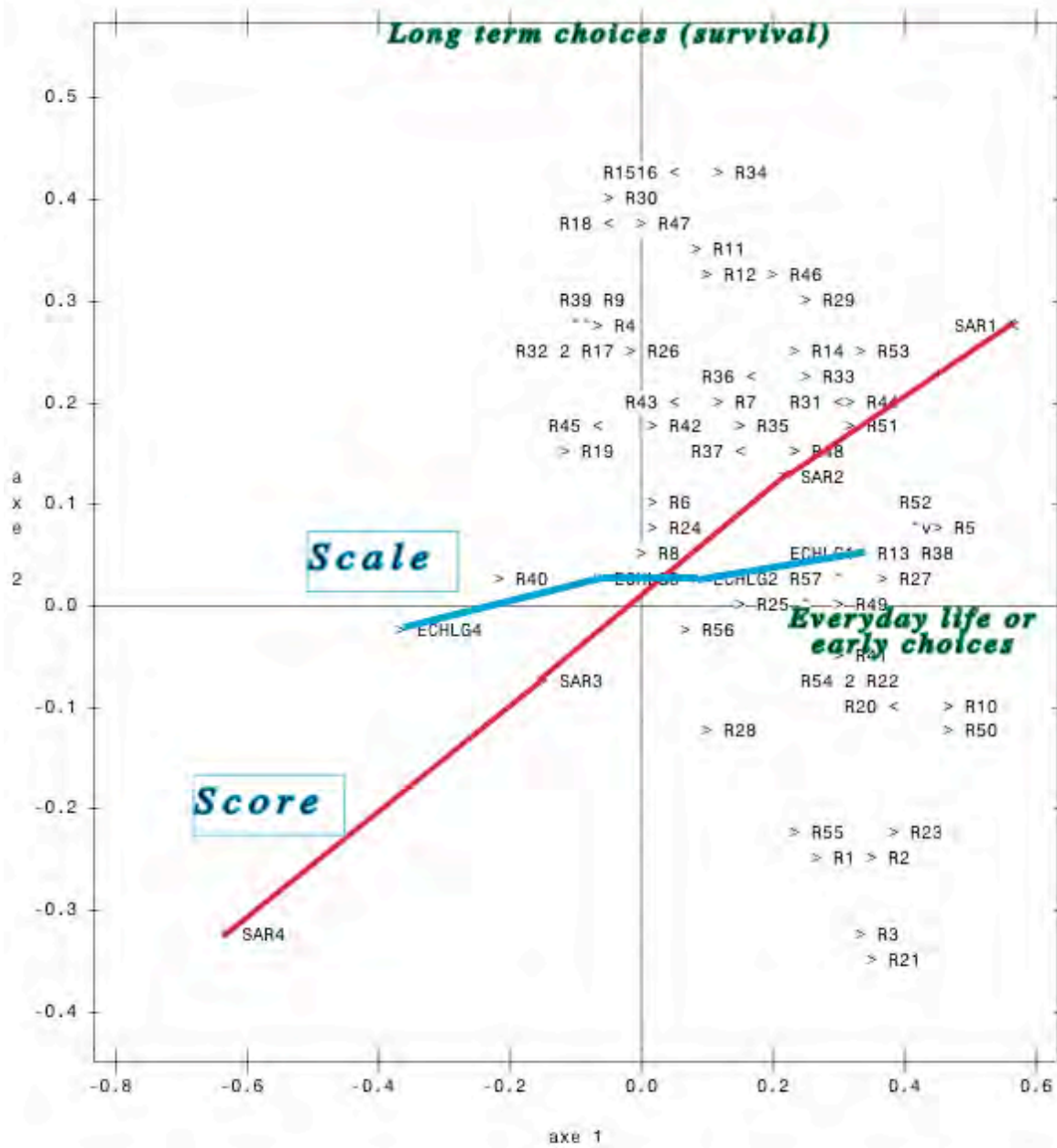
Source: Patrimoine 1998, Insee-Delta survey.

Graph 2 : Risk and Time Preference and wealth accumulation



Source: Patrimoine 1998, Insee-Delta survey.

Graph 3: Principal component analysis : risk
(2 first axes)



Source: Patrimoine 1998, Insee-Delta survey.