Network for Studies on Pensions, Aging and Retirement

Netspar THESES

Ineke Bijlsma

Changes in Household Behavior after a Housing Wealth Shock

MSc Thesis 2013-024

Changes in Household Behavior after a Housing Wealth Shock

Master Thesis for the Master *Econometrics and Mathematical Economics*
School of Economics and Management

Abstract

This paper investigates the effects of a sudden housing wealth shock on household savings. We deviate from other research with a model where housing wealth is set up as a risky investment, as it is speculated this was the behavior of the Dutch before the financial crisis. We then estimate the effect of housing wealth on net savings using household survey panel data and a system generalized method of moments estimator. For young homeowners the results are in line with the general predictions of our model. For older homeowners there is an interaction term with age that makes the effect of housing wealth on savings insignificant. However, the predicted marginal effect of younger homeowners is as large as 0.6, far greater than most literature, and 0.21 without accounting for age. Future research could expand the model to include age effects and added detail to increase both the realism of the model as well as the validity of the results, allowing for a better economic interpretation of these findings.

Ineke Bijlsma (ANR: 756376)

Under supervision from Dr. M.J. Bijlsma
of the CPB Netherlands Bureau for Economic Policy Analysis
and Prof. A.H.O. van Soest of Tilburg University

December 11, 2013

Contents

1 Introduction

The current financial crisis, started by the bursting of the U.S. housing bubble, still leaves its mark on the worldwide economy today. Relative low growths and volatile markets are still common, especially in Europe where some countries were and still are in danger of not being able to repay and refinance government debt. While careful celebrations are made as the eurozone shows signs of recovery, this is mainly due to the positive economic growth of some countries such as Germany. Other major European countries still struggle, including the Netherlands. Despite the recovery of its neighbors, it is entering its third recession since the start of the crisis.

Due to their housing economy, it seems the Dutch have been hit harder than its neighbors by the bursting of the housing bubble. Rising property values since the nineties combined with economic growth and mortgage interest being fully tax deductible gave rise to a distortion of the housing market, with national prices rising 11% annually on average during the peak years 1996-2001 and media house prices rising by about 80% around that same period (Statistics Netherlands). The bursting of the housing bubble brought the prices down to 23% between 2007 and 2012, and the resulting shock to housing wealth left the Dutch with high mortgages and debt-to-income ratio of around 250%, far higher than the average of about 100% of the euro area (Eurostat). Without the financial security of their house to depend on during the ongoing crisis, household behavior changed as consumption dropped to accommodate to the situation.

The effects of housing wealth on consumption and savings are not a new field, and most studies do find results where lower housing wealth signifies lowered consumption levels. However, depending on the research the elasticity between the two differs a lot, both including as low as 0.02 and as high as 1.7. The reason for this discrepancy is likely a result of the difference in theoretical approach, suggesting the approach taken is highly important for the found elasticity of housing wealth on consumption. Some examples: Yao and Zhang (2005) use housing as an ability to borrow by collateralizing the house, and estimate the difference between renting and owning a house with the use of panel data. Campbell and Cocco (2007) estimate the response of household consumption to house prices on a pseudo-panel by using unexpected housing wealth shocks per period in the regression analysis. Bostic et al. (2009) use an unique dataset with a lot of information and estimate consumption over several log-variables per three years, Carroll et al. (2011) use a 'sticky expectations' model for households with aggregate data, and the working paper of Hurd et al. (2013) incorporates a differences-in-differences estimator for the recession period. Mian et al. (2013) also use data from the recession period, but without using panel data and focusing more on the role of debts.

This paper adds a different approach to housing wealth compared to the ones above. One theory about the housing market crash is that households invested as much as possible in housing in the period before the crisis due to the risky but prospective housing wealth growths it offered. Concurrently, saving would then be equal to a safe investment with lower but secure returns. This approach of treating the two as types of investments is different from the methods described above, but might be a better fit for the situation in the Netherlands than current methods. If the results coincide with the model, it might give new theoretical insights on how to treat housing wealth during such times. With a more accurate way to predict household behavior, current policies can be improved.

We build a toy model modeling the basic theoretical approach, forgoing complexity for a better understanding of the effects of this approach due to the lack of complex interactions. In chapter 2 we list the assumptions and specify the consumer optimization problem for both the full model and a simplified version. We solve these in chapter 3, as well as using comparative statics to fully understand the model and its behavior.

After forming the necessary hypotheses from the model, we detail the empirical analysis in chapter 4. We use data from the DNB Household Survey using a dynamic panel data regression. Some of the advantages of this survey are that it is updated annually and includes questions about psychological concepts such as future discounting and risk aversion. However, one of the disadvantages is that after the necessary transformations the panel is highly unbalanced. Furthermore, the rich are overrepresented, although we try to remedy this with using weights. We also explain in depth the choice for a system GMM estimator and the subsequent results.

In the last chapter we interpret our results and draw conclusions about the effect of housing wealth on savings and the fit of our model. We also discuss possible improvements for the model for further research. The appendix includes mathematical calculations, survey questions and other background material.

2 The Model

2.1 Assumptions

We use a simple model of consumption smoothing to study how savings respond to house price movements. To capture the effects of a house price shock in such a model, it includes a situation where at a time t a price drop just occurred in the housing price. A household has to adjust consumption C_t and savings S_t taking this into account. For this we need at least one period before and after to get a non-trivial problem. First there needs to be a past point $t-1$ where a household invested a nonnegative amount H_{t-1} in housing, not knowing about the negative shock that will occur. The second is a future period $t+1$ such that there is a future to save up for. These are the minimal assumptions we need to account for.

As we wish to make the model as simple as possible, we only take into account the periods outlined above. Starting at a time $t = 0$, a household has an initial monetary endowment y_0 from which he chooses to invest an amount $H_0 \geq 0$ in his house. The remainder of y_0 is available for $t = 1$. Then before $t = 1$ a shock to the housing price occurs. With probability p the housing price rises to a level $\overline{\alpha}H_0$ or lowers with probability $(1-p)$ to level $\underline{\alpha}H_0$, where $\overline{\alpha}, \underline{\alpha}$ are constants with $\overline{\alpha} > 1 > \underline{\alpha} > 0$. With this information at $t = 1$ the household chooses between consumption and saving. The last period $t = 2$ is the end of the life cycle of the household. The house is sold for its price and together with his savings rS_1, with $r \geq 1$, all is spent on consumption.

Other important assumptions are that we do not take into account renting, moving, or otherwise being able to sell a house at the period $t = 1$, as well as utility derived from housing as a consumption good. Furthermore, as $t = 2$ signals the end of the life cycle, all debts must be paid off. While housing is a risky investment, one advantage is the result of this investment already being known at $t = 1$. This means consumers can change their choices between consumption and saving depending on the outcome of the investment. We summarize all above assumptions in Figure 1, with the money streams per period.

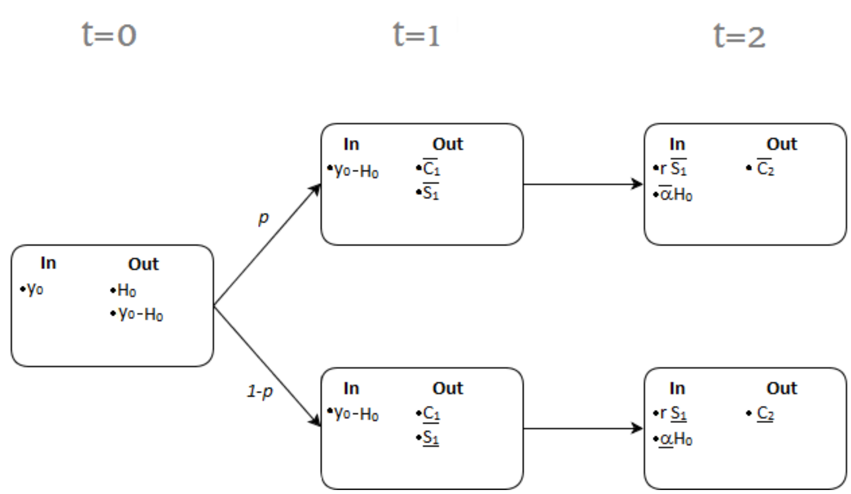


Figure 1: An overview of the money streams in each period t. C_t is the consumption level, S_t is the money set aside for saving, and H_t is the money spent on housing.

Additional assumptions are consumption always being nonnegative, there being a discount rate $0 < \delta < 1$ such that consumption in the present is encouraged, and the utility function satisfying $u'(.) > 0$ and $u''(.) < 0$ such that it gives diminishing returns of consumption at higher levels. Based on this, we define the lifetime utility per household as

$$U = u(C_1) + \delta u(C_2)$$

Here $u(C_t)$ is the utility function per period. For this function, we consider both the natural logarithm and the whole family of isoelastic utility functions with risk aversion parameter η due to their constant relative risk aversion. More information about the properties of these utility functions can be found in Appendix A.

As the consumer tries to maximize his expected utility under the restrictions placed on him, the goal is to calculate the choices he makes based on given consumer-specific parameters. One advantage of using the two-period model is that the derivative of the expected utility function is analytically solvable, while a three-period is not.

2.2 Consumer optimization problem

To get a feeling for the model, we first look at a subgame of the full model starting at $t = 1$. As a subgame, the solution should be the same as the solution in the full model at $t = 1$. Define y_1 as all wealth left from $t = 0$ and redefine h_0 as the current housing price after the shock. We then get the set of equations

$$\begin{aligned} \max_{C_1, C_2, S_1} \ & E_1(U) = u(C_1) + \delta u(C_2) \\ \text{s.t. } & C_1 = y_1 - S_1, \\ & C_2 = rS_1 + h_0, \\ & C_1, C_2 \geq 0 \end{aligned} \tag{1}$$

Here the function to be maximized is the expected utility function in period 1 while the restrictions are based on the balancing of the money streams of the first and second period.

For the full model we include $t = 0$. As we are dealing with the expected utility from $t = 0$ instead of $t = 1$, we must average out utility over the high and low housing price situation. Furthermore, substitute in (1) the following: $y_1 = y_0 - H_0$, $h_0 = \overline{\alpha} H_0$ for the high house pricing situation and $h_0 = \underline{\alpha} H_0$ for the low house pricing situation. The full consumer optimization problem is then given by

$$\begin{aligned} \max_{\overline{C_1}, \overline{C_2}, \overline{S_1}, \underline{C_1}, \underline{C_2}, \underline{S_1}, H_0} \ & E_0(U) = p \left[u(\overline{C_1}) + \delta u(\overline{C_2}) \right] + (1-p) \left[u(\underline{C_1}) + \delta u(\underline{C_2}) \right] \\ \text{s.t. } & \overline{C_1} = y_0 - H_0 - \overline{S_1}, \\ & \overline{C_2} = r\overline{S_1} + \overline{\alpha} H_0, \\ & \underline{C_1} = y_0 - H_0 - \underline{S_1}, \\ & \underline{C_2} = r\underline{S_1} + \underline{\alpha} H_0, \\ & \overline{C_1}, \overline{C_2}, \underline{C_1}, \underline{C_2}, H_0 \geq 0. \end{aligned} \tag{2}$$

3 Solution of the Model

3.1 Solution of the simple model

For the simple model, we test two additional assumptions the model can have in tandem, so up to four different versions of the same model. The first rule is a borrowing constraint that prevents a consumer from borrowing money; in other words, an added restriction $S_1 \geq 0$. The second rule is about the shape of the utility function: to take the natural logarithm or the more general family of isoelastic utility functions with variable risk aversion parameter η (as detailed in Appendix A). In subsection 3.3, we show comparative graphs for all the four models side by side. In this section, we work out one of the models analytically. We assume we have variable η and no borrowing constraints present. The other three models are be solved in the same way.

Start off with assuming we are given values for parameters δ, r, y_1, h_0 and η. The first two restrictions of (1) give C_1 and C_2 in terms of S_1, so substitute this in the function to be maximized. The conditions $C_1, C_2 \geq 0$ give lower and upper bounds for S_1:

$$\max_{\frac{-h_0}{r} \leq S_1 \leq y_1} \frac{(y_1 - S_1)^{1-\eta} - 1}{1-\eta} + \delta \frac{(rS_1 + h_0)^{1-\eta} - 1}{1-\eta} \tag{3}$$

The first-order condition of (3) is

$$-(y_1 - S_1^*)^{-\eta} + \delta r (rS_1^* + h_0)^{-\eta} = 0 \tag{4}$$

If we solve (4) as in Appendix B to calculate the optimal value of S_1 and fill it into the other equations, we get

$$\begin{aligned} S_1(\delta, r, y_1, h_0, \eta) &= \frac{\gamma y_1 - h_0}{r + \gamma} \text{ with } \gamma = \sqrt[\eta]{\delta r} \\ C_1(\delta, r, y_1, h_0, \eta) &= y_1 - S_1(\delta, r, y_0, h_0, \eta) \\ C_2(\delta, r, y_1, h_0, \eta) &= rS_1(\delta, r, y_0, h_0, \eta) + h_0 \\ E_1(U(\delta, r, y_1, h_0, \eta)) &= \frac{C_1(\delta, r, y_1, h_0, \eta)^{1-\eta} - 1}{1-\eta} + \delta \frac{C_2(\delta, r, y_1, h_0, \eta)^{1-\eta} - 1}{1-\eta} \end{aligned} \tag{5}$$

3.2 Solution of the full model

Again, assume we are given the necessary parameters $\delta, r, y_0, \overline{\alpha}, \underline{\alpha}$ and p. Unlike the simple model, the full model is not analytically solvable for some η, so we fix $\eta = 1$. Start with the solutions of the subgame last chapter and substitute in $y_1 = y_0 - H_0$, $h_0 = \overline{\alpha} H_0$ for the high house pricing situation and $h_0 = \underline{\alpha} H_0$ for the low house pricing situation, as done before when defining the consumer optimization problem. (5) then becomes

$$
\begin{aligned}
\overline{S_1}(H_0,\delta,r,y_0,\overline{\alpha}) &= \frac{\delta r(y_0-H_0)-\overline{\alpha}H_0}{r(1+\delta)}\\
\overline{C_1}(H_0,\delta,r,y_0,\overline{\alpha}) &= y_0-H_0-\overline{S_1}(H_0,\delta,r,y_0,\overline{\alpha})\\
\overline{C_2}(H_0,\delta,r,y_0,\overline{\alpha}) &= r\overline{S_1}(H_0,\delta,r,y_0,\overline{\alpha})+\overline{\alpha}H_0\\
\\
\underline{S_1}(H_0,\delta,r,y_0,\underline{\alpha}) &= \frac{\delta r(y_0-H_0)-\underline{\alpha}H_0}{r(1+\delta)}\\
\underline{C_1}(H_0,\delta,r,y_0,\underline{\alpha}) &= y_0-H_0-\underline{S_1}(H_0,\delta,r,y_0,\underline{\alpha})\\
\underline{C_2}(H_0,\delta,r,y_0,\underline{\alpha}) &= r\underline{S_1}(H_0,\delta,r,y_0,\underline{\alpha})+\underline{\alpha}H_0
\end{aligned}
\tag{6}
$$

If we substitute the values of C_1 and C_2 from (6) back into the utility function of the full model, we get as the optimal value of H_0:

$$
\begin{aligned}
\max_{0\le H_0\le y_0} E_0(U) = p&\left[\ln\left(y_0-H_0-\overline{S_1}(H_0)\right)+\delta\ln\left(r\overline{S_1}(H_0)+\overline{\alpha}H_0\right)\right]\\
&+(1-p)\left[\ln\left(y_0-H_0-\underline{S_1}(H_0)\right)+\delta\ln\left(r\underline{S_1}(H_0)+\underline{\alpha}H_0\right)\right]
\end{aligned}
\tag{7}
$$

If we calculate the first-order condition of this function as in Appendix C, we get

$$
H_0 = \begin{cases} 0 & \text{if } H_0^* < 0 \\ y_0 & \text{if } H_0^* > y_0 \\ H_0^* & \text{otherwise} \end{cases}
$$

$$
\text{Where } H_0^* = \frac{-\,ry_0(\overline{\alpha}p+\underline{\alpha}(1-p)-r)}{(\overline{\alpha}-r)(\underline{\alpha}-r)}
$$

3.3 Comparative statics

Our model depends on the variables we assign to the consumer, and a good understanding of the model is needed to properly form our hypotheses. To examine the influence of a variable on the model, we fix the other values and plot the graph over the variable to be examined. If we do this for all variables, we should be able to draw conclusions of the behavior of the model and form hypotheses for the empirical analysis.

For the examination, it is important to choose the fixed variables carefully. We want to avoid corner solutions where the consumer optimization problem becomes irrelevant. One example would be if we choose the fixed variables such that returns on savings in period 2 would be greater than the expected returns on housing wealth; in that case, no one would buy a house which is both unrealistic and does not properly show the effects of the change in variables. For each model, we fix our variables in a point that is not too unrealistic, shows clear changes in the graphs and starts in none of the corner solutions.

For the simple model, we choose the fixed point

$$\{\delta = 0.9, r = 1.1, h_0 = 3, y_1 = 5, \eta = 0.5 \text{ (when applicable)}\}$$

We research the effect on C_1, C_2, S_1 and U_t per variable across the four different models discussed earlier. Figure D2 in Appendix D shows us the results.

We notice some results that are visible in all models. The ratio between C_1 and C_2 changes with δ and r, but it remains the same for h_0 and y_0 unless income cannot be properly balanced over the two periods due to borrowing constrains. We also see that if η is very small, the intertemporal rate of substitution is very high, meaning it is better to consume immediately even with a very small negative incentive.

Between the models with and without borrowing, the biggest changes are in low δ and low y_1. With low δ, it is far better to consume as much in the first period, but the consumer is limited by his income y_1. This also happens if y_1 is low in comparison to h_0. Therefore, borrowing constraints punish those with low income. As it is reasonable to borrow in such circumstances and the amount borrowed is already limited by a fraction of housing wealth, we do not plan to include borrowing constraints. Lastly, we notice the models with lower η give different slopes for change in δ and r. Otherwise, the results are fairly similar.

We also use comparative statics for the full model, and choose a fixed point in the same way as the simple model. This time, we pick the point

$$\{r = 1.1, \delta = 0.8, p = 0.5, \overline{\alpha} = 2, \underline{\alpha} = 0.5, y_0 = 100\}$$

Figure D3 in Appendix D then shows us the results for the full model.

The graphs for change in p, y_0 and difference in $\overline{\alpha}$ and $\underline{\alpha}$ are as expected. High p gives a higher expected value for the returns of the investment in H_0 which therefore leads to an increase in H_0 by households. For change in $\overline{\alpha}$ and $\underline{\alpha}$ there are diminishing returns due to risk aversion, even as the expected value for investment in H_0 gets higher. The result of y_0 was assumed in the model due to constant relative risk aversion.

For r, the model behaves the same as the simple model until it is not optimal anymore to make any investment in H_0. For $H_0 = 0$, r leaves the equation of C_1 and S_1, which means only C_2 increases. For δ, the model behaves exactly the same as the simple model, as H_0 does not depend on δ as we saw when solving the model. Intuitively we can explain this that for high δ the investment in housing is used to finance the consumption in $t = 2$, and for low δ it is used to pay off debts accumulated in period 1.

As one of our goals is to see whether the results from the empirical analysis are in line with predictions of our model, we set up a number of hypotheses based on the solutions and graphs of the model. In the empirical analysis, we will choose net savings S_1. This is because utility is not measurable, consumption is not included in the data, and starting housing wealth depends on a number of variables that are hard to measure like p.

Furthermore, $\overline{\alpha}H_0$ is irrelevant here as our data is from the financial crisis, focusing on the low-term situation where the housing prices drop to $\underline{\alpha}H_0$. p is not measurable and r is constant across individuals; meaning its effect will be incorporated in the time dummies included in the regression. We are then left with variables y_0, $\underline{\alpha}H_0$, δ and η.

The original endowment y_0, is not identifiable in empirical analysis. We approximate the term $(y_0 - H_0)$ with the savings of last period S_{t-1}. Likewise, $\underline{\alpha}$ is not directly identifiable, but the term $\underline{\alpha}H_0$ is equal to the current housing price H_t, possibly needed to be lagged if tests indicate this due to the assumption the household needs to know whether they are in the good or the bad situation. For δ and η we construct indices from survey questions to approximate the effect. The four variables to include in the regression will therefore be S_{t-1}, H_t, δ and η. We'd get a regression like

$$S_{it} = \beta_1 S_{i,t-1} + \beta_2 H_{it} + \beta_3 \delta_{it} + \beta_4 \eta_{it} + \beta_5 X_{it} + u_{it} \tag{8}$$

Where X_{it} is a vector of control variables and u_{it} is the error term assumed to have mean zero.

The most basic condition the hypotheses need to satisfy is that the signs of the coefficients of the variables are correct. Since $y_0 - H_0$ has a positive influence on S_1 in the formula of our model, S_{t-1} should too, and therefore β_1 should be positive. Likewise, the coefficient of H_t, β_2, should have a negative effect on net savings and therefore be negative. From the graphs it is likely the coefficient for δ, β_3, is positive, only being insignificant when borrowing constrains are present. For η, we expect its coefficient β_4 to be insignificant: In the simple model we saw η only affects savings when close to zero. In the real world such a risk-neutral approach would be very rare (Janeček, 2004).

We also make further predictions on the sizes of the coefficients. In (6) we have, after substitution, the following equation for S_t:

$$S_t = \frac{\delta r \; S_{t-1} - H_t}{r(1+\delta)} \tag{9}$$

It is unclear whether δr will be greater or smaller than 1 in (9), so therefore we cannot make any predictions which coefficient, β_1 or $|\beta_2|$, would be greater. However, assuming the assumptions $r > 1$ and $(1+\delta) > 1$ are true, both β_1 or $|\beta_2|$ should at least be less than 1.

For the size of β_3, it would require some speculation. If we take a look at the graph of change in δ of Figure D3, it appears from extrapolating that β_3 should roughly be 30. Of course, the effect of δ on S_t is a lot more complex and depends also on the other values, so it is hard to give upper and lower bounds like we did for β_1 and β_2. We can only speculate δ should likely have a positive significant effect, assuming there are no budget constrains.

Summarized, we have the four following hypotheses:

$$\begin{array}{lcrcl} H_I & : & 0 < & \beta_1 & < 1 \\ H_{II} & : & -1 < & \beta_2 & < 0 \\ H_{III} & : & 0 < & \beta_3 & \\ H_{IV} & : & & \beta_4 & = 0 \end{array} \tag{10}$$

These hypotheses reflect the minimum validity needed to accept the basics of our model, and we shall check them in the results section 4.3.

One disadvantage of only looking at level terms is that we do not directly measure how the change in housing wealth causes a change in savings (and consumption). One possibility to introduce the term ΔH_t to (9) is to take first differences such that $\Delta S_t = S_t - S_{t-1}$. The result of this transformation would be

$$\Delta S_t = \frac{\delta r \ \Delta S_{t-1} - \Delta H_t}{r(1+\delta)} \tag{11}$$

By replacing S_t, S_{t-1} and H_t with $\Delta S_t, \Delta S_{t-1}$ and ΔH_t in (8), we expect to draw similar conclusions as set up in (10) and get better result for the effects of a housing wealth shock. However, the difference terms do have some problems on their own. First of all, some data is lost due to the increased number of lags needed to compute the results. Second, by looking at the differences, the original levels of the variables are lost, which is information on its own about the choices consumers made in the (distant) past regarding savings and housing. As such, the accuracy of the results is likely to be lower than when we use the level terms. Third, if people like to keep their savings constant if there are no housing wealth shocks, $\Delta S_t = \Delta S_{t-1} = 0$ when $\Delta H_t = 0$, censoring all data outside large housing wealth drops and causing even further inaccuracy. Therefore, we choose to show the results for both (9) and (12), keeping in mind the strength and weaknesses of both regressions.

4 Empirical Analysis

4.1 Data

We take data from the DNB Household survey (DNB HHS) over the period 2000 to 2012. The DNB HHS is a Dutch annual survey starting from 1993. It covers a number of subjects about general household information, work, housing, income, health, assets, debts and perceptions about economic and psychological concepts, including discount rates and risk aversion. Furthermore, as an online survey participants can answer questions whenever they want and with all the documents they need, there is an decreased number of missing or inaccurate values.

We construct panel data by taking the identification number of each household as the individual dimension and the year of the survey as the time dimension. As the DNB HHS has an over-representation of high income levels, we construct probability weights per income category per year. We use these weights in the regression such that the observations reflect the income distribution of the Dutch population. We sum per household the assets, total debt and net financial income. For housing wealth we take the expectation of the sale price instead of the official value used for taxation. We also convert all financial numbers to 2005 euros ($\times 1000$).

We first construct the net financial wealth of a household by summing all assets and then subtracting all liabilities. However, the net financial wealth of a household includes the residence, and we want to measure the effect of lowered housing wealth on the rest of wealth. We therefore construct a variable we will call net savings (S_t), which we consider the net financial wealth minus the self-identified housing wealth (H_t) of the residence. Note we the term net savings does not solely contain all assets, but also has all liabilities subtracted from it. This is because we consider paying off mortgages and other liabilities a form of saving up.

For δ and η, as they cannot be measured directly, we instead opt to create an index of those two variables based on the responses to questions about discount rates and risk aversion. For the time-variant individual characteristics we take net income, age, partner present, number of children and long-term health problems. We furthermore construct dummies for each type of region, urbanization, education, and the current year. For a more detailed overview of the used survey questions and construction of the data, see Appendix E. Furthermore, we use a number of robustness checks in section 4.4 to reject a number of different specifications.

From this dataset we select only those who are head of their household and homeowners, leaving us with 5444 households for our data. However, due to the assumptions of our model, we need to drop or censor some of the data. First, we drop all observations of homeowners who are 65 and older since they are not representative for the time $t = 1$ in our model; after all, the assumption of a future period to save up is not certain for this age group. Second, we censor housing wealth (and thus also net savings) for those with housing wealth over a million euro. It is unclear whether or not it is an answering mistake due to the scale one needs to answer in, and looking over the data gives reasonable doubt these type of errors are present. Lastly, we remove observations of housing wealth where it made a jump upwards of at least 500 thousand euros compared to the other values for the same reason. Because of the missing data, this selection is a highly unbalanced panel: only 4.30% has filled in data for each year, while over 25% of the households has only data for one year.

In Figure C a histogram of each of the (unweighted) regressors is given on the years 2007 and 2012, representing the distributions of just before and after the peak of the crisis. Housing wealth seems to follow a positive-skewed distribution like the chi-squared distribution if we smooth out the tail (due to self-reported amounts often being rounded). Net savings doesn't seem to follow an easily identifiable probability distribution.

Since the discount rate and risk aversion are indices and can only take set values, the histogram bars sometimes take zero with values in between seemingly overrepresented. However, if one would smooth out these values, we'd likely get a normal distribution as drawn into the histogram. Keep in mind the distribution of income can be different in the two years due to the unbalancedness of the panel, and therefore these histograms are only an estimate to make sure our data is reliable.

For housing wealth the mode in 2007 is slightly right from the €200,000 mark, while in 2012 it is left of it. If we round the values to 210 and 180, this would mean a difference of about 15% of the two modes. Net savings increases in dispersion in 2012 compared to 2007; it seems some people get more into debt while others save up more instead. By including the two indices and the lagged variable of wealth, we hope to be able to separate these two kinds of people.

Where in 2007 the mode of the index of δ is around 0.65, in 2012 the mode seems to have shifted to 0.7, which means the financial crisis seems to have inspired people to think more about the future. At the other hand, the distribution of the index of η is unchanged.

Table D1 in Appendix D shows the summary statistics of the pooled data. Net savings is negative on average, but this is expected since mortgages are included while housing wealth is not. If we include housing wealth, the net financial wealth is shown to be about 215 thousand euros on average. Also note that our constructed indices have a low standard deviation, which might signify a small deviation being enough to cause a big shift in extra savings.

Table D2 shows the distribution of housing wealth per year while Table D3 shows net savings. The housing price mean starts dropping from 2011 onwards likely because it takes some time for people to accept how bad the housing market has become. However, the data of official prices has its own problems due to uneven lengths of measurement times giving uneven results, as well as only being available from 2003 onwards. Furthermore, it can be argued the perception of housing wealth is more important to saving and consumption behavior than the true value. We do a robustness check in section 4.4 using this value to support this claim.

The most important result of Table D3 is that there is an unnatural distribution in the year 2000 when it comes to net savings. As such, we should censor all data of net savings for that year. A possible reason for this is that in 2000 the move to online communication and other transitions was still going on, which might have had an influence on the results of the survey. Excluding that, the dispersion of net savings increases over time as we also saw in the histograms.

In the end, we have data with 8403 observations over 2058 households who have no missing data over the main regressors. Which data can be further used depends on the estimator and the variables included in the regression. For each regression, it is therefore reported how many observations and households are used.

4.2 Regression model

We set up a regression based on the theoretical model. We limit ourselves to the variables $S_t, S_{t-1}H_t, \delta$ and η as discussed in section 3.3. We also include both time dummies, time-variant and (unobserved) time-invariant household characteristics as control variables. The regression equation for the level terms is then equal to

$$S_{it} = \beta_0 + \beta_1 S_{i,t-1} + \beta_2 H_{it} + \beta_3 \delta_{it} + \beta_4 \eta_{it} + \beta_5 \varphi_t + \beta_6 X_{it} + \alpha_i + u_{it} \quad (12)$$

Where S_{it} is net savings of this period and the dependent variable, $S_{i,t-1}$ is net savings of last period, H_{it} is housing wealth, δ_{it} is the index of the discount rate, η_{it} is the index of risk aversion, φ_t are the year dummies, X_{it} are time-variant household characteristics, α_i is the time-invariant individual effect and u_{it} is the error term with mean zero. When we wish to measure the effect on differences, we use

$$\Delta S_{it} = \beta_0 + \beta_1 \Delta S_{i,t-1} + \beta_2 \Delta H_{it} + \beta_3 \delta_{it} + \beta_4 \eta_{it} + \beta_5 \varphi_t + \beta_6 X_{it} + \alpha_i + u_{it} \quad (13)$$

Because we include dummy variables in (12) and (13), both in year dummies and household characteristics, we need to drop one dummy variable per category to avoid perfect multicollinearity. We choose to drop the year 2000 from the year dummies and the first answer of each survey question for the separate household dummies. This results in the constant β_0 now including the effect of the dropped dummy variables and becoming both time- and household variant. Since the constant and controls are of no interest in our case, we do not report it.

If we would ignore it is panel data and apply pooled OLS regression, the estimator might suffer from omitted variables bias due to correlation between the error terms of the same households as well as ignoring the term α_i. By using panel-clustered standard errors one can take into account the potential correlation between the error terms, but the effect of α_i on the dependent variable is still not properly being taken into account. For (12), this correlation is positive and the lagged variable also bears a positive effect on net savings, so therefore it is biased upwards. For (13), ΔS_{it} and $\Delta S_{i,t-1}$ will be negatively correlated with one another.

If we use fixed effects or first differences as a panel data estimator, the time-invariant effects will be transformed out of the equation. However, there will be another form of bias, Nickell bias (Nickell, 1981), due to the lagged variable term now being correlated with part of the error term. This bias is of the order $\frac{1}{T}$ for $N \to \infty$, and the direction of the bias depends on T. As our example is rather unbalanced and has relatively small T, fixed effects is not a good estimator. The same problem also affects the random effects model.

We can solve this problem by acknowledging the dynamic panel data structure and using second and further lags as instruments for the lagged variable, either as differences or as levels. For i.i.d. error terms the lagged instruments will be highly correlated with the (difference of the) dependent variable, while the error terms will be uncorrelated due to the difference in time. This process is called the Anderson-Hsiao estimator (Anderson and Hsiao, 1981). However, it is not a very efficient estimator due to the use of a single lag as instrument.

We choose to use a two-step feasible system GMM estimator. GMM is designed for panels with a low number of time periods compared to the number of households, which is the case in our dataset. With it, we can generate seperate instrument(s) for each lag and time period instrumented. Therefore, one can eliminate the trade-off between lag length and sample length and include as many lagged variable instruments as necessary, increasing efficiency (Roodman, 2009). We estimate the lagged variable with even deeper lagged instruments of both levels and differences and use the other variables, year dummies and household characteristics as additional exogenous instruments of their level term. We also correct the downward bias for the two-step approach, which is needed to deal with potential heteroskedasticity.

Furthermore, instead of first differencing, we opt for an forward orthogonal deviation (FOD) transformation. FOD subtracts the average of all available future observations from the term of the current period, and then multiplies it with a scale factor chosen such that if the original error terms are i.i.d, then so are the transformed ones. The (i, t)th observation after the FOD transformation of a variable x, given at least one future observation, would be

$$FOD.x_{it} = \lambda(x_{it} - \frac{1}{T'} \sum_{\bar{t}=t+1}^{T} x_{i\bar{t}})$$

Here λ is the scale factor and T' is the number of future non-missing periods. On balanced panels this gives numerically identical coefficient estimates, but FOD helps to preserve sample size in panels with gaps (Arellano and Bover 1995). As the fixed effects estimator estimates the coefficient of x with the coefficient of the transformed $FE.x$, in GMM we estimate the coefficient of x on the transformed variable $FOD.x$.

We also report the number of instruments and the Hansen J p-statistic in the regression results. The Hansen J-test from Hansen (1982) is a test that checks the necessary orthogonality conditions of GMM by means of a constructed J-statistic. Under a well-specified model with valid moment conditions it behaves like a chi-squared random variable. Rejection of this null hypothesis indicates endogenous instruments or other forms of misspecification. However, a high number of instruments makes the Hansen J-test weak. Therefore, should we be satisfied with the results, we run the GMM estimator three times: The first includes for all endogenous variables all their lags as instruments that are at least two periods removed from the original. The second one uses those from 2 and 3 periods ago, while the third only uses instruments lagged 2 periods away from the endogenous variable. This way, we can check the robustness of the results when we decrease the number of instruments. We also make sure then there is no AR(2+) present and that the Difference-in-Hansen tests for both the lagged variable instruments subset and the other instrument subset also do not reject the null hypothesis of the Hansen J-test when the more general test does not reject, both at the 10% level (both not reported).

By using the results to see whether the hypotheses in (10) are rejected or not for both (12) and (13), we can make a statement over if the results support our theoretical model and its findings. Furthermore, we also show the results for OLS and FE to confirm our suspicions outlined in this section and to increase belief in the final results.

4.3 Results

Table 1 gives the results of our two base regressions. Specification (1) and (3) are the results of the regression over levels and first differences respectively, estimated with OLS, FE and GMM.

If we compare the results in column (1) with the hypotheses set up in paragraph 3.3, the results are not in line with the predictions made. The results for OLS and GMM do not reject H_I, H_{III} and H_{IV}, but does reject H_{II}. The FE estimates in turn satisfy H_{II} at the 5% level, but reject all other hypotheses. One possibility for the difference between the model and the results of the regression could be due the inclusion of household characteristics. However, if we exclude them as in column (2), the conclusions we draw remain the same. As such, the model is not yet representative of consumer behavior when looking at level terms.

For column (3), the main hypothesis H_{II} is satisfied for all estimators. ΔS_{t-1} is negative in OLS and FE estimation, but insignificant for GMM. As OLS and FE are both biased, the result from GMM should be the most accurate one. This means we reject all hypotheses aside from H_{II}, which is in line with one of the speculations of possible weaknesses when we set up the regression in section 3.3.

The main discrepancy in results between the two regressions is that the difference regression implies there is a negative effect of housing wealth on net savings, yet no such result is there for the level regression. Furthermore, while age is insignificant for the difference regression, it is definitely significant for the level regression. This all implies age effects are missing in the model, where the time left in the model period $t = 1$ significantly influences savings behavior of the consumer.

To test out this conjecture, we add an interaction term $(H \times Age)_{it}$ to (12). Preliminary results show household characteristics causes the null hypothesis of the Hansen J test to be severely weakened. As we also saw this happening for the difference equation in Table 1, we remove them. The new equation, before transformation with fixed effects or FOD, is then equal to

$$S_{it} = \beta_{0_{it}} + \beta_1 S_{i,t-1} + \beta_2 H_{i,t} + \beta_3 \delta_{it} + \beta_4 \eta_{it} + \beta_5 \ (H \times Age)_{i,t} + \beta_6 \varphi_t + \alpha_i + u_{it} \qquad (14)$$

With the year dummy 2000 dropped. Likewise, for the difference equation, we replace $S_{it}, S_{i,t-1}$, H_{it} and $(H \times Age)_{it}$ with their first differences to get a similar result for (13). The results of the new equations are located in Table 2. The three columns of the GMM specification are equal to the three GMM specifications outlined in section 4.2, where we decrease the number of instruments used. Columns (6) and (12) use the same estimator as columns (5) and (11), but with included household characteristics to compare the results.

Table 1: Base regression results.

	Dependent variable: S_t						Dependent variable: ΔS_t					
	(1)			(2)			(3)			(4)		
	OLS	FE	GMM	OLS	FE	GMM	OLS	FE	GMM	OLS	FE	GMM
S_{t-1}	0.760***	0.233	0.597***	0.780***	0.222	0.620***						
	(0.0182)	(0.190)	(0.0383)	(0.0158)	(0.175)	(0.0414)						
H_t	0.00865	-0.215*	-0.0335	0.0429	-0.254**	0.0184						
	(0.0449)	(0.0926)	(0.0547)	(0.0435)	(0.0868)	(0.0417)						
ΔS_{t-1}							-0.227***	-0.292***	-0.184	-0.235***	-0.310***	-0.181
							(0.0554)	(0.0505)	(0.141)	(0.0543)	(0.0523)	(0.147)
ΔH_t							-0.226*	-0.231*	-0.202*	-0.247**	-0.267**	-0.210*
							(0.0880)	(0.0918)	(0.101)	(0.0885)	(0.0946)	(0.0967)
Index δ	58.73**	35.98	82.73***	69.98***	55.07	94.66***	2.147	3.542	3.868	8.458	32.81	6.326
	(20.67)	(37.92)	(22.17)	(20.76)	(34.64)	(19.60)	(26.69)	(41.89)	(20.27)	(24.04)	(39.27)	(17.07)
Index η	29.84	22.38	6.259	34.56*	16.09	21.60	49.43*	76.81*	17.21	44.64*	70.32*	16.87
	(15.61)	(24.27)	(17.46)	(14.84)	(22.38)	(17.09)	(19.79)	(37.74)	(15.73)	(19.47)	(33.45)	(16.09)
Age	1.328***	-0.0307	1.982***				0.0473	1.239	0.166			
	(0.207)	(2.193)	(0.357)				(0.271)	(8.416)	(0.203)			
Net income	0.424	-0.388	0.599*				-0.263	-1.118	-0.135			
	(0.358)	(0.668)	(0.266)				(0.288)	(0.783)	(0.154)			
No. of observations	3949	3949	3949	4383	4383	4383	2971	2971	2971	3267	3267	3267
No. of households		1028	1028		1116	1116		776	776		851	851
Instrument count			87			68			76			57
Hansen J p-value			0.0757			0.0193			0.0537			0.177

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: All values are in 2005 euros ($\times 1000$). Specifications (1) and (3) include household characteristics X_{it} in the regression (not reported aside from age and income) while (2) and (4) exclude household characteristics. Year dummies and the constant term are included but not reported. The GMM estimator transforms the variables over forward orthogonal deviations, where S_{t-1} or ΔS_{t-1} are instrumented with lags. All other independent variables are used as instruments in GMM. The Hansen J-test checks for the null hypothesis the system is 'valid' with exogenous instruments. The standard errors shown in parentheses are either corrected with household-clustered standard errors (OLS, FE) or Windmeijer's finite-sample correction for the two-step covariance matrix (GMM).

Table 2: Adjusted regression results.

	Dependent variable: S_t						Dependent variable: ΔS_t					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	OLS	FE	GMM	GMM	GMM	GMM (+X_{it})	OLS	FE	GMM	GMM	GMM	GMM (+X_{it})
S_{t-1}	0.755***	0.220	0.584***	0.605***	0.579***	0.573***						
	(0.0150)	(0.177)	(0.0413)	(0.0397)	(0.0687)	(0.0678)						
H_t	-0.316***	-0.531	-0.477***	-0.432***	-0.473***	-0.642**						
	(0.0573)	(0.353)	(0.0878)	(0.0792)	(0.101)	(0.212)						
$(H \times Age)_t$	0.00640***	0.00575	0.00873***	0.00814***	0.00897***	0.0119**						
	(0.00123)	(0.00668)	(0.00131)	(0.00124)	(0.00169)	(0.00450)						
ΔS_{t-1}							-0.237***	-0.312***	-0.185	-0.00196	-0.0535	-0.0844**
							(0.0546)	(0.0519)	(0.148)	(0.0251)	(0.0404)	(0.0294)
ΔH_t							-1.069**	-1.051**	-1.002**	-0.874*	-0.778*	-0.848*
							(0.336)	(0.380)	(0.345)	(0.362)	(0.350)	(0.373)
$\Delta(H \times Age)_t$							0.0167**	0.0157*	0.0160*	0.0129	0.0118	0.0132
							(0.00630)	(0.00719)	(0.00629)	(0.00671)	(0.00642)	(0.00688)
Index δ	71.38***	56.84	94.45***	96.41***	105.3***	100.8***	7.605	32.51	5.527	19.29	17.12	2.329
	(20.22)	(34.63)	(19.82)	(19.81)	(22.82)	(25.76)	(24.14)	(39.16)	(16.93)	(18.28)	(18.90)	(22.17)
Index η	21.63	16.98	7.750	15.24	7.347	11.52	48.72*	73.59*	20.91	39.94*	42.01**	44.77**
	(14.82)	(22.56)	(17.58)	(18.08)	(18.15)	(20.10)	(19.44)	(33.45)	(15.48)	(15.59)	(15.39)	(16.78)
Age						-0.494						-0.0785
						(0.968)						(0.224)
Net income						0.543**						0.0175
						(0.174)						(0.194)
No. of observations	4383	4383	4383	4383	4383	3949	3267	3267	3267	3267	3267	2971
No. of households		1116	1116	1116	1116	1028		851	851	851	851	776
Instrument count			69	41	33	52			58	37	30	49
Hansen J p-value			0.0302	0.0337	0.324	0.0147			0.214	0.248	0.334	0.269

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: All values are in 2005 euros (×1000). Year dummies and the constant term are included in all specifications but not reported. Specifications (6) and (12) include household characteristics X_{it} in the regression (not reported aside from age and income). The GMM estimator estimates the variables over forward orthogonal deviations where H_t and S_{t-1} are instrumented with lags. Level values of the indices and year dummies are used as instruments in GMM. The GMM specifications differ in the amount of lags used as instrumental variables for the lagged variable. The Hansen J-test checks for the null hypothesis the system is 'valid' with exogenous instruments; we give its p-value to see whether it rejects the null hypothesis. The standard errors shown in parentheses are either corrected with household-clustered standard errors (OLS, FE) or Windmeijer's finite-sample correction for the two-step covariance matrix (GMM).

For the level regression in Table 2, there is a clear change now that we have included the interaction term with age. Both H_t and $(H \times Age)_t$ are significant for all specifications but the fixed effects estimator, confirming out conjecture regarding age interaction effects being present. As $(H \times Age)_t$ is positive, this means a younger homeowner saves up more if a drop in housing wealth occurs, while older homeowners are less affected. Therefore, for this group all four hypotheses are satisfied, and we conclude the model and its predictions are in line with the results.

If homeowners are old enough, however, the marginal effect of a change in housing wealth will be zero, as can be seen from Table 3. H_{II} is not satisfied and the predictions of the model are not in line with the results. There can be multiple reasons for this result. One might be that households change their saving behavior at certain ages, reacting differently to lowered housing wealth. Another reason could be that due to the housing wealth increase before the crisis, there is no need to cut consumption and save up for the future as the investment was still profitable in the long term expectations. General housing wealth increase in age could also mean the interaction term between age and housing wealth is more of an indicator for high housing wealth in general, which would lead to completely different conclusions. As our model and predictions are not equipped to test these speculations, we leave it as it is.

If we include household characteristics as in column (6), we notice that the results seem to be in the confidence interval of the specification without household characteristic in (5), but they violate the null hypothesis of the Hansen J-test. As such, the preferred specification for the level terms is column 5 as it is the only specification where the p-value is higher than 10%.

For the difference equation, both housing wealth and the interaction effect with age are significant in the strongest GMM specification in column (9), though only at the 5% level. The other variables are insignificant, which is the same result as we saw earlier in Table 1. Here too we include one of the estimates with household characteristics. While the null hypothesis of the Hansen J-test is not violated like for the level equation, it is still weakened. In the end, our preferred specification for the difference estimator is (9) as it includes the most instruments without rejecting the necessary conditions.

Table 3: Marginal effect of housing wealth on net savings for fixed ages of the specifications (5) and (9) in Table 2.

Age	25	35	45	55
H_t on S_t	-0.248***	-0.158**	-0.0688	0.0210
	(0.0627)	(0.0496)	(0.0396)	(0.0352)
ΔH_t on ΔS_t	-0.603**	-0.443**	-0.284**	-0.124
	(0.195)	(0.140)	(0.0956)	(0.0804)

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4.4 Robustness checks

In the previous sections we defined the regression and determined the effect of housing wealth on savings for both difference and level terms. We used difference-in-Hansen tests to test for exogeneity within the instrument set, a two-step approach to deal with heteroskedasticity within individuals, the Windmeijer correction for the two-step standard errors and AR(2+) tests to deal with autocorrelation. However, there could still be a number of specifications that would better fit the model and assumptions to be checked. For this reason, we do a number of robustness checks.

First of all, we address the problem of the existence of a unit root. If a series has an unit root, any shocks to the time series are permanent, whereas with no unit root it returns to the trend eventually. The problem with our dataset is that it is both unbalanced and has a rather low number of periods compared to the number of households we test, which invalidates most of the tests for unit roots. On the other hand, the low number of time periods compared to the large number of households also means false results due to non-stationarity should be limited, as we are mainly interested in the short-term drop of net savings. Therefore, we speculate any effects of a possible unit root present are minimal enough to be ignored.

The second important issue are the level and difference equations we specified. Especially for the latter, we listed its weaknesses in section 3.3 and were confirmed in the results in section 4.3. To sidestep those issues, we could have replaced H_t with ΔH_t in (8), such that we get

$$S_{it} = \beta_1 S_{i,t-1} + \beta_2 \Delta H_{it} + \beta_3 \delta_{it} + \beta_4 \eta_{it} + \beta_5 X_{it} + u_{it} \tag{15}$$

If we make the same changes to this equation as to the difference equation in section 4.3, we would get the following results

Table 4: Elasticity of housing wealth on net savings for fixed ages of the specifications (5) and (8) in Table 2.

	(1)	(2)	(3)	(4)	(5)
	OLS	FE	GMM	GMM	GMM
S_{t-1}	0.783***	0.231	0.631***	0.650***	0.628***
	(0.0138)	(0.175)	(0.0446)	(0.0402)	(0.0656)
ΔH_t	-0.834***	-0.536*	-0.728**	-0.677**	-0.675**
	(0.244)	(0.220)	(0.251)	(0.256)	(0.243)
$\Delta(H \times Age)_t$	0.0127**	0.00787	0.0112*	0.0105*	0.0106*
	(0.00465)	(0.00442)	(0.00483)	(0.00484)	(0.00460)
Index δ	70.30**	55.41	88.46***	88.70***	103.1***
	(21.63)	(34.92)	(19.12)	(19.06)	(23.07)
Index η	29.61*	10.28	22.48	30.72	23.44
	(14.84)	(23.20)	(16.91)	(16.65)	(16.79)
No. of observations	4383	4383	4383	4383	4383
No. of households		1116	1116	1116	1116
Instrument count			69	41	33
Hansen J p-value			0.0222	0.0252	0.321

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results are significant and column (5) has a Hansen J p-statistic above the 10% level. However, there are two reasons why we have chosen not to draw any conclusions over these results. The first is a minor one in that equation (9) cannot be transformed such that we use ΔH_t to estimate S_t instead of H_t, which would mean the current model is not properly tested to make the appropriate hypotheses. The second reason is that when adding household characteristics as controls, both ΔH_t and $\Delta(H \times Age)_t$ become insignificant. While we removed the household characteristics from the final regression, we did this because it weakened the GMM estimator, but they led to the same conclusions about the validity of the model. We therefore conclude this specification is not reliable enough to draw conclusions from.

We also check several different specifications regarding lagged variables. We try to replace the first lag of net savings with the second lag, housing wealth with the first, second or third lag, and the indices with their first lags. Using the first lag of housing wealth violates the Hansen J p-test for all but the highest amount of instruments, and as the test is known to be weak for high instrument counts we should not use this specification. However, using the second lag for the level equation gives a significant result, as can be seen in Table 5.

Table 5: Regression results with dependent variable S_t for H_{t-2}.

	(1)	(2)	(3)	(4)	(5)
	OLS	FE	GMM	GMM	GMM
S_{t-1}	0.714***	0.248	0.519***	0.550***	0.566***
	(0.0288)	(0.167)	(0.0659)	(0.0653)	(0.0636)
H_{t-2}	-0.337***	-0.231	-0.981**	-0.788*	-0.702*
	(0.0753)	(0.408)	(0.326)	(0.376)	(0.280)
Index δ	72.48**	35.53	79.53*	99.95**	111.4***
	(22.47)	(37.76)	(39.77)	(32.36)	(30.75)
Index η	33.42	24.89	-3.465	12.41	10.83
	(17.74)	(27.12)	(30.83)	(28.92)	(26.31)
$H_{t-2} \times Age$	0.00687***	0.00229	0.0197**	0.0155*	0.0135**
	(0.00176)	(0.00763)	(0.00739)	(0.00666)	(0.00524)
Instrument count			175	91	67
Hansen J p-value			0.352	0.275	0.270

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: All values are in 2005 euros (×1000). Year dummies and the constant term are included but not reported. The GMM estimator estimates the variables over forward orthogonal deviations where H_{t-2} and S_{t-1} are instrumented with lags. Level values of the indices and year dummies are used as instruments in GMM. The GMM specifications difer in the amount of lagged instruments used. The Hansen J-test checks for the null hypothesis the system is 'valid' with exogenous instruments; we give its p-value to see whether it rejects the null hypothesis. The standard errors shown in parentheses are either corrected with household-clustered standard errors (OLS, FE) or Windmeijer's finite-sample correction for the two-step covariance matrix (GMM).

However, like with using equation (15) there are some issues with this specification. First of all, using a similar substitution for the difference equation yields insignificance for the new variable, and as such there seems to be no direct effect of ΔH_{t-2} on ΔS_t. Secondly, for the economic implication it seems unusual households would make decisions based on their perceived value of their residence of two years ago instead of taking the present value. As such, we prefer the old specification. All other replacements we specified earlier lead to the new specification variable being insignificantly different from zero. We also try to add the new lags above separately to the regression instead of replacing the old ones, to similar effect.

We also try different specifications for the variables. We construct a number of different indices for δ and η by taking subsets of the currently used questions, for example by dropping the questions that are not answered on the 1 to 7 scale. The estimates of the effects of these indices show no significant differences compared to the current estimates. We also try the official housing wealth value instead of the self-reported one, but it has an insignificant effect on net savings, likely due to its measurement problems. Furthermore, we both try specifications where we exclude all real estate wealth from S_t, include all real estate wealth in H_t, or both. The results become insignificant for housing wealth (and at times other variables), so it seems non-residential real estate does behave like wealth and not like main housing wealth.

Another check we do is taking the natural logarithm over housing wealth. While it is insignificant when we regress with level terms, for the difference equation we do find an significant effect whilst the interaction term with age becomes insignificant. However, as it is only significant at the 5% level with very high standard errors we address it as a bad fit for our chosen regression.

Table 6: Regression results of $\Delta \ln H_t$ for a alternate specification of Table 2.

	(1)	(2)	(3)	(4)	(5)
	OLS	FE	GMM	GMM	GMM
$\Delta \ln H_t$	-75.54*	-58.57	-69.98*	-65.23*	-58.58*
	(31.37)	(34.01)	(32.87)	(27.53)	(28.17)
Instrument count			58	37	30
Hansen J p-value			0.183	0.255	0.343

Standard errors in parentheses
* $p < 0.05$

Note: Only $\Delta \ln H_t$ is reported.

We also rerun the regressions with both uncensored data aside from age, or with censored data but including all ages. For the first specification without outside instruments, the coefficients for H_t and $H_t \times Age$ have so much variance they become statistically insignificant from zero. If we include the household characteristics as instruments they do fall within the 95% confidence interval of those coefficients. Likewise, the same happens with the second specification.

The last test we run is construct age dummies per 10 years with the first one starting from age 15 and let them interact with the second lag of housing wealth, instead of the interaction term $H_t \times Age$. The results of this regression are similar to what we found in Table 2, only do we also find the effect of housing wealth insignificant on the 15-24 age group. Considering the average finances of this group, it is likely they don't have the budget to make significant changes. They also number low enough that dropping them from the data makes minimal difference in the results and well within the 95% confidence interval.

5 Conclusion

The goal of this paper was to set up a model where housing wealth is defined as a risky investment and to analyze the results of this model compared to empirical analysis. Based on the solutions of the consumer optimization problem and the comparative statics, we set up a number of hypotheses to test for the empirical analysis. We then used data from the DNB Household Survey to estimate the effects of housing wealth on net savings in a GMM framework. We also did the same taking first differences for savings and housing wealth. The base regression without first differences rejected the hypothesis concerning housing wealth, indicating the base model is not yet representative of consumer behavior.

By adding an interaction term of housing wealth with age to the regression, we found the results of younger homeowners are in line with the predicted behavior, both when using levels and using first differences. For older homeowners of the age of 50-55 the results show a marginal effect to be zero and thereby violating one of the model hypotheses.

For those of 25 years of age we find the marginal effect of net savings with respect to housing wealth to be around -0.6. If we assume the increase in wealth comes from a decrease in consumption, this would be equal to a marginal effect on consumption of 0.6 with respect to change in housing wealth. Most literature finds a marginal propensity to consume below 0.1, which means a value of 0.6 is very high in comparison. Even if we would leave out age effects, from Table 1 we find marginal effect of 0.21 which is still greater than current predictions.

The economic interpretation for this result is made more difficult by the fact we have not proven why the interaction term needs to be included in the regression. For example, the high marginal effect might be due to the high loan-to-income ratio of the Netherlands, as Mian et al. (2013) showed leveraged households have an higher marginal propensity to consume. But we could also conclude older homeowners might suffer from a mortality risk, which is included in for example Hurd et al. (2013), and the interaction term with age is just a way to estimate its effect.

An updated model that includes age effects is needed to properly investigate housing wealth effects on savings and consumption, testing all speculations to find the correct one. With better specifications, one could test for the reason for the age interaction effects and its effect on the model. This together with added detail should increase the validity and reliability of the results, and proper economic conclusions could be drawn. Afterwards further research could be done whether the model also fits similar countries or time periods.

References

[1] Anderson, T.W. and Cheng Hsiao. (1981). *Estimation of dynamic models with error components.* Journal of the American Statistical Association, 76 (375), 589-606.

[2] Arellano, M. and Bover. O. (1995). *Another look at the instrumental variable estimation of error-components models.* Journal of Econometrics, 68 (1), 29-51.

[3] Bostic, R., Gabriel, S. and Painter, G. (2009). *Housing Wealth, Financial Wealth, and Consumption: New Evidence from Micro Data.* Regional Science and Urban Economics, 39 (1), 79-89.

[4] Bover, O. (2005). *Wealth Effects on Consumption: Microeconometric Estimates from the Spanish Survey of Household Finances.* Unpublished.

[5] Carrol, C.D., Otsuka, M. and Slacalek, J. (2011). *How Large are Housing and Financial Wealth Effects? A New Approach.* Journal of Money, Credit and Banking, 43 (1), 55-79.

[6] Campbell, J.Y. and Cocco, J.F. (2004). *How Do House Prices Affect Consumption? Evidence from Micro Data.* Journal of Monetary Economics, 54 (3), 591-621.

[7] Deaton, A. (1992). *Understanding Consumption.* Oxford: Clarendon Press.

[8] Hansen, L.P. (1982). *Large Sample Properties of Generalized Method of Moments Estimators.* Econometrica, 50 (4), 1029-1054

[9] `http://cdata3.uvt.nl/dhs/users/login` (2013). *DHS data access*, November 11, 2013.

[10] Angrisani, M., Hurd, M.D. and Rohwedder, S. (2013). *Changes in Spending in the Great Recession Following Stock and Housing Wealth Losses.* Netspar working paper.

[11] Janeček, K. (2004). *What is a realistic aversion to risk for real-world individual investors?* Carnegie Mellon University Working paper.

[12] Mian, A., Rao, K. and Sufi, A. (2013). *Household Balance Sheets, Consumption and the Economic Slump.* Chicago Booth Research Paper No. 13-42; Fama-Miller Working Paper.

[13] Nickell, S. (1981). *Biases in Dynamic Models with Fixed Effects.* Econometrica, 49 (6), 1417-1426.

[14] Roodman, D. (2009). *How to do xtabond2: An introduction to difference and system GMM in Stata.* Stata Journal, 9(1), 86-136.

[15] Roodman, D. (2009). *A Note on the Theme of Too Many Instruments.* Oxford Bulletin of Economics and Statistics, 71(1), 135158.

[16] Teppa, F. and Vis, C. (2012). *The CentERpanel and the DNB Household Survey: Methodological Aspects.* DNB occasional studies, 10 (4).

[17] Yao, R. and Zhang, H.H. (2005). *Optimal Consumption and Portfolio Choices with Risky Housing and Borrowing Constraints.* Journal of Money, Credit and Banking, 43 (1), 197-239.

Appendix

A Properties of the chosen utility functions

The two functions we will use as utility functions are the natural logarithm, and the more general family of isoelastic utility functions

$$u(c) = \begin{cases} \frac{c^{1-\eta}-1}{1-\eta} & \text{if } \eta \geq 0,\ \eta \neq 1 \\ \ln(c) & \text{if } \eta = 1 \end{cases}$$

Where c is the level of consumption, ln is the natural logarithm and η a constant. η is related to two properties of the utility function: the risk aversion of consumers and the intertemporal elasticity of substitution. We will go in more detail about these two properties.

Risk aversion is a property of consumers which means they gain lower utility from an uncertain payoff than a certain one with the same expected gain. To show this in action, take a look at time period 2 in our model, where a consumer has invested a nonzero amount H_0 in housing wealth and spends it on consumption C_2. Then, using the natural logarithm as our utility function for this example, the expected utility gain from investment in the house will be

$$E\left(U(H_0)\right) \quad = p\ln\left(\overline{\alpha}H_0\right) + (1-p)\ln\left(\underline{\alpha}H_0\right)$$

This is less than the utility of the expected profit $U\left(E(H_0)\right) = \ln\left(p\overline{\alpha} + (1-p)\underline{\alpha}\right)H_0$, as can be proved by using Jensens inequality for concave functions. Therefore, the utility gained from H_0 will be smaller if $\overline{\alpha}$ and $\underline{\alpha}$ differ more, assuming the expected value of the monetary profit remains the same.

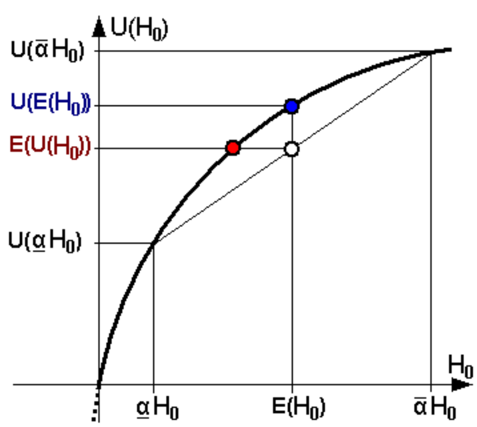


Figure A1: An overview of Jensen's inequality.

The fine property of the isoelastic utility functions is that the Arrow-Pratt-De Finetti measure of relative risk aversion (RRA), is constant. The RRA per period t is defined as

$$RRA_t(C_t) = C_t A(C_t)$$

where C_t is the level of consumption in period t and $A(C_t) = -\frac{U''(C_t)}{U'(C_t)}$ (the absolute risk aversion) is the curvature of $U(C_t)$.

For $U = \frac{C_t^{1-\eta}-1}{1-\eta}$ we have $A(C_t) = -\frac{-\eta\ C_t^{-\eta-1}}{C_t^{-\eta}} = \frac{\eta}{C_t}$, which means $RRA = \eta$ for all time periods t. As the relative risk aversion for C_t is not actually dependent on C_t itself, this means the relative level of consumption, the fraction of wealth spent on consumption in period t, is unaffected by initial wealth. If the absolute risk aversion would be constant, then the absolute level of consumption would be unaffected by initial wealth.

In other words, if we scale initial wealth y_0 with an coefficient, the levels of consumption, savings and housing corresponding with y_0 are scaled with the same coefficient.

Because of $RRA = \eta$, we have that $\eta = 0$ corresponds to risk neutrality while $\eta \to \infty$ corresponds to infinite risk aversion.

The second property is the intertemporal elasticity of substitution, which is a measure of responsiveness to the incentive $(r - \delta)$ to save up. The formula for this is

$$\frac{\frac{\partial}{\partial t}C_t}{C_t} = -\frac{-\ U'(C_t)}{C_t U''(C_t)}\,(r-\delta)$$

For $U = \frac{C_t^{1-\eta}-1}{1-\eta}$, we get

$$\frac{\frac{\partial}{\partial t}C_t}{C_t} = -\frac{C_t^{-\eta}}{\eta\ C_t^{-\eta}}(r-\delta) = \frac{r-\delta}{\eta}$$

If η is small, consumers will react a lot to changes in the incentive $(r-\delta)$. If η is large, consumers do not change their current consumption much even if the incentive to do so is very large.

One final note: both the risk aversion and intertemporal rate of substitution do not depend on the level of consumption but instead depend on the constant η. This means they cannot be separated as one cannot be changed without changing the other as well.

B Solving the simple model

Continuing from the first-order condition:

$$\begin{aligned} -(y_1 - S_1^*)^{-\eta} + \delta r(rS_1^* + h_0)^{-\eta} &= 0 \\ \delta r(rS_1^* + h_0)^{-\eta} &= (y_1 - S_1^*)^{-\eta} \\ \delta r(y_1 - S_1^*)^{\eta} &= (rS_1^* + h_0)^{\eta} \end{aligned}$$

Let $\gamma = \sqrt[\eta]{\delta r}$. We then get

$$\gamma(y_1 - S_1^*) = rS_1^* + h_0$$

$$\frac{\gamma y_1 - h_0}{r + \gamma} = S_1^*$$

Furthermore, the second-order condition gives

$$-\eta\left((y_1 - S_1^*)^{-\eta-1} + \delta r^2(rS_1^* + h_0)^{-\eta-1}\right) \quad < 0$$

which means S_1^* is indeed a maximum.

The last thing we need to check before we conclude $S_1 = S_1^*$ is whether S_1^* lies within the restricted domain $\left[\frac{-h_0}{r}, y_1\right]$ in all cases.

If $S_1^* < \frac{-h_0}{r}$, the following equation should hold:

$$\begin{array}{lcr} \dfrac{\gamma y_1 - h_0}{r + \gamma} & < & \dfrac{-h_0}{r} \\ r(\gamma y_1 - h_0) & < & -(r + \gamma)h_0 \\ ry_1 & < & -h_0 \end{array}$$

As $h_0, y_1 \geq 0$ this does not hold.

If $S_1^* > y_1$, the following equation should hold:

$$\begin{array}{lcr} \dfrac{\gamma y_1 - h_0}{r + \gamma} & > & y_1 \\ \gamma y_1 - h_0 & > & (r + \gamma)y_1 \\ -h_0 & > & ry_1 \end{array}$$

Again, it doesn't hold. Therefore, we end up with $S_1(\delta, r, y_1, h_0, \eta) = S_1^*(\delta, r, y_1, h_0, \eta)$. To get the values for C_1 and C_2, use the restrictions from before:

$$\begin{aligned} C_1 &= y_1 - S_1 \\ C_2 &= rS_1 + h_0 \end{aligned}$$

Then fill in S_1, C_1 and C_2 to get the maximum expected utility.

C Solving the full model

If we continue with the simplified set of equations

$$\begin{aligned}
&\max_{0 \leq H_0 \leq y_0} E_0(U) \\
&\quad = p\left[\ln\left(y_0 - H_0 - \overline{S_1}(H_0)\right) + \delta \ln\left(r\overline{S_1}(H_0) + \overline{\alpha} H_0\right)\right] \\
&\qquad + (1-p)\left[\ln\left(y_0 - H_0 - \underline{S_1}(H_0)\right) + \delta \ln\left(r\underline{S_1}(H_0) + \underline{\alpha} H_0\right)\right] \\
&\quad = p\left[\ln\left(y_0 - H_0 - \frac{\delta r(y_0 - H_0) - \overline{\alpha} H_0}{r(1+\delta)}\right) + \delta \ln\left(\frac{\delta r(y_0 - H_0) - \overline{\alpha} H_0}{1+\delta} + \overline{\alpha} H_0\right)\right] \\
&\qquad + (1-p)\left[\ln\left(y_0 - H_0 - \frac{\delta r(y_0 - H_0) - \underline{\alpha} H_0}{r(1+\delta)}\right) + \delta \ln\left(\frac{\delta r(y_0 - H_0) - \underline{\alpha} H_0}{1+\delta} + \underline{\alpha} H_0\right)\right] \\
&\quad = p\left[\ln\left(\frac{r(y_0 - H_0) - \overline{\alpha} H_0}{r(1+\delta)}\right) + \delta \ln\left(\frac{\delta r^2(y_0 - H_0) + \delta r\overline{\alpha} H_0}{r(1+\delta)}\right)\right] \\
&\qquad + (1-p)\left[\ln\left(\frac{r(y_0 - H_0) - \underline{\alpha} H_0}{r(1+\delta)}\right) + \delta \ln\left(\frac{\delta r^2(y_0 - H_0) + \delta r\underline{\alpha} H_0}{r(1+\delta)}\right)\right]
\end{aligned}$$

Remove constants

$$\begin{aligned}
&= p\left[\ln\left(r(y_0 - H_0) - \overline{\alpha} H_0\right) + \delta \ln\left(\delta r^2(y_0 - H_0) + \delta r\overline{\alpha} H_0\right)\right] \\
&\quad + (1-p)\left[\ln\left(r(y_0 - H_0) - \underline{\alpha} H_0\right) + \delta \ln\left(\delta r^2(y_0 - H_0) + \delta r\underline{\alpha} H_0\right)\right] \\
&= p\left[\ln\left(ry_0 - (\overline{\alpha} + r)H_0\right) + \delta \ln\left(\delta r^2 y_0 + \delta r(\overline{\alpha} - r)H_0\right)\right] \\
&\quad + (1-p)\left[\ln\left(ry_0 - (\underline{\alpha} + r)H_0\right) + \delta \ln\left(\delta r^2 y_0 + \delta r(\underline{\alpha} - r)H_0\right)\right]
\end{aligned}$$

If we take the derivative of this and set it to zero, we get the first-order condition:

$$\frac{(1+\delta)\big(-H_0(\overline{\alpha} - r)(\underline{\alpha} - r) + ry_0(-\underline{\alpha}(1-p) - \overline{\alpha} p + r)\big)}{\big((r - \overline{\alpha})H_0 - ry_0\big)\big((\underline{\alpha} - r)H_0 + ry_0\big)} = 0$$

To solve this first-order condition, we first check whether the denominator is zero anywhere on the domain. This happens when $(r - \overline{\alpha})H_0 - ry_0 = 0$ or $(\underline{\alpha} - r)H_0 + ry_0 = 0$, which is equal to

$$(r-\overline{\alpha})H_0 = ry_0 \quad \text{or} \quad (r-\underline{\alpha})H_0 = ry_0$$

If we include the assumption made in section 2.1 that $(r-\overline{\alpha}) < (r-\underline{\alpha}) < r$, both cases are excluded over the domain. Therefore, we can multiply both sides with the denominator to get the simplified expression:

$$-H_0^*(\overline{\alpha}-r)(\underline{\alpha}-r) + ry_0\big(-\underline{\alpha}(1-p) - \overline{\alpha}p + r\big) = 0$$

If we solve this over H_0^*, we'd get

$$H_0^* = \frac{-ry_0\big(\overline{\alpha}p + \underline{\alpha}(1-p) - r\big)}{(\overline{\alpha}-r)(\underline{\alpha}-r)}$$

And so, the full expression of H_0 is

$$H_0 = \begin{cases} 0 & \text{if } H_0^* \leq 0 \\ y_0 & \text{if } H_0^* \geq y_0 \\ H_0^* & \text{otherwise} \end{cases}$$

D Tables and figures

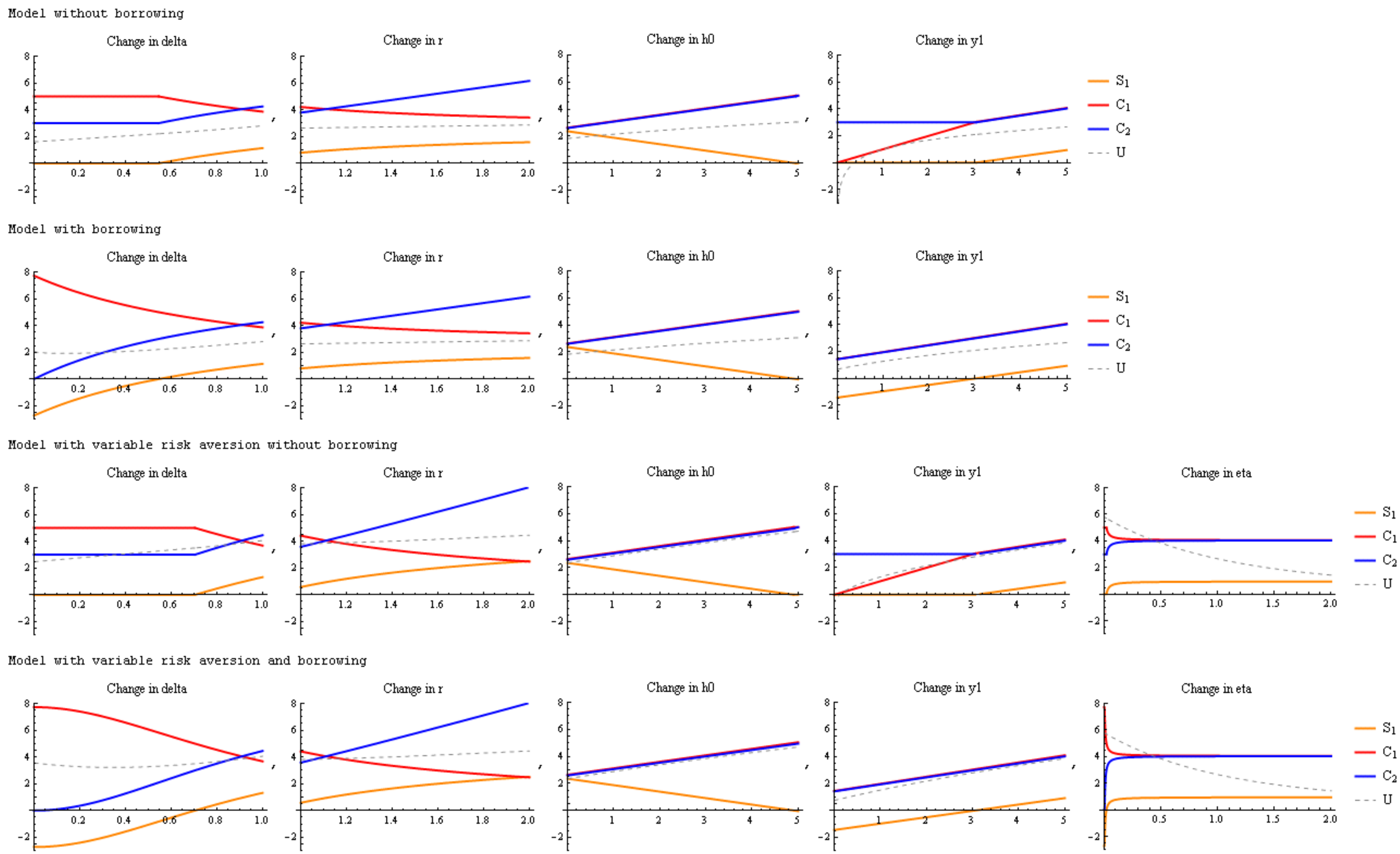


Figure D2: Comparative statics for four different models

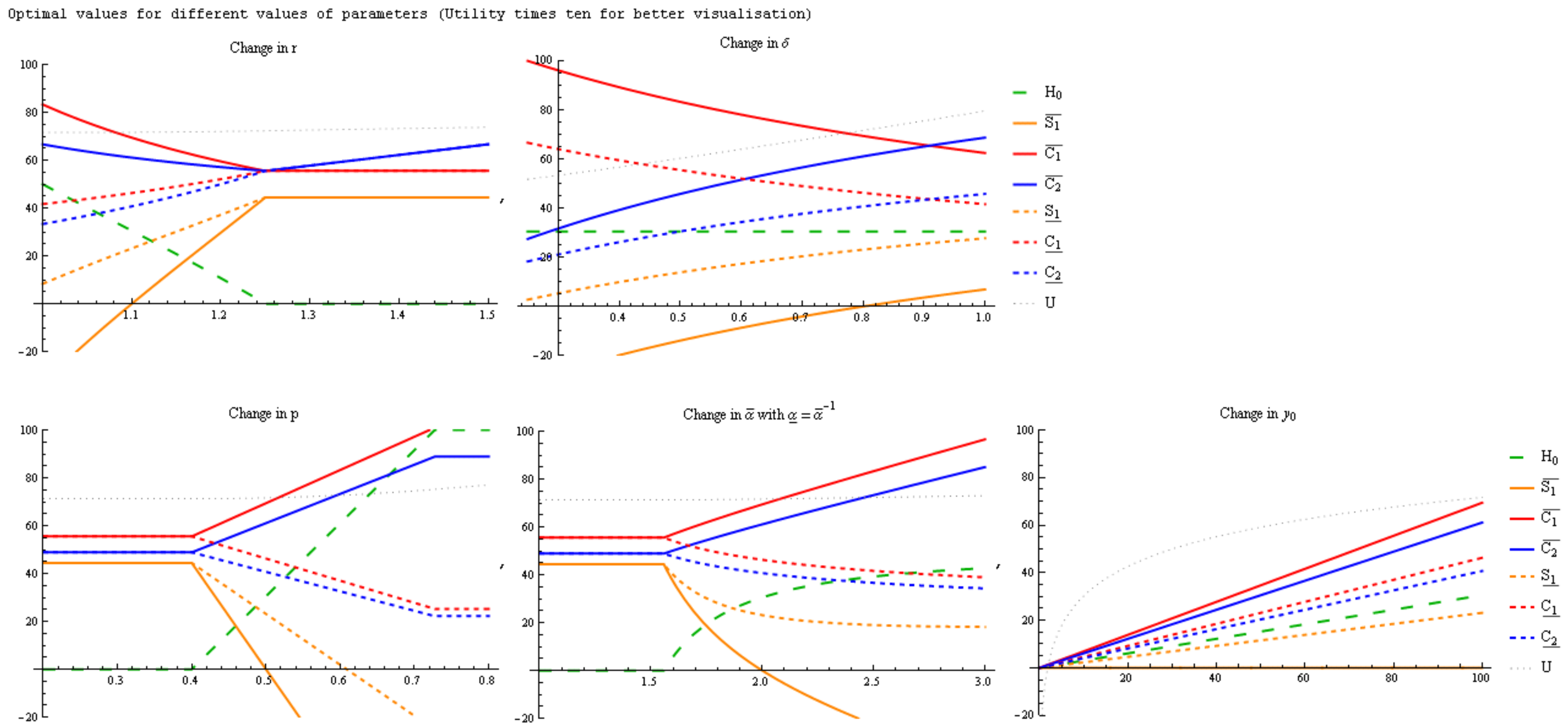
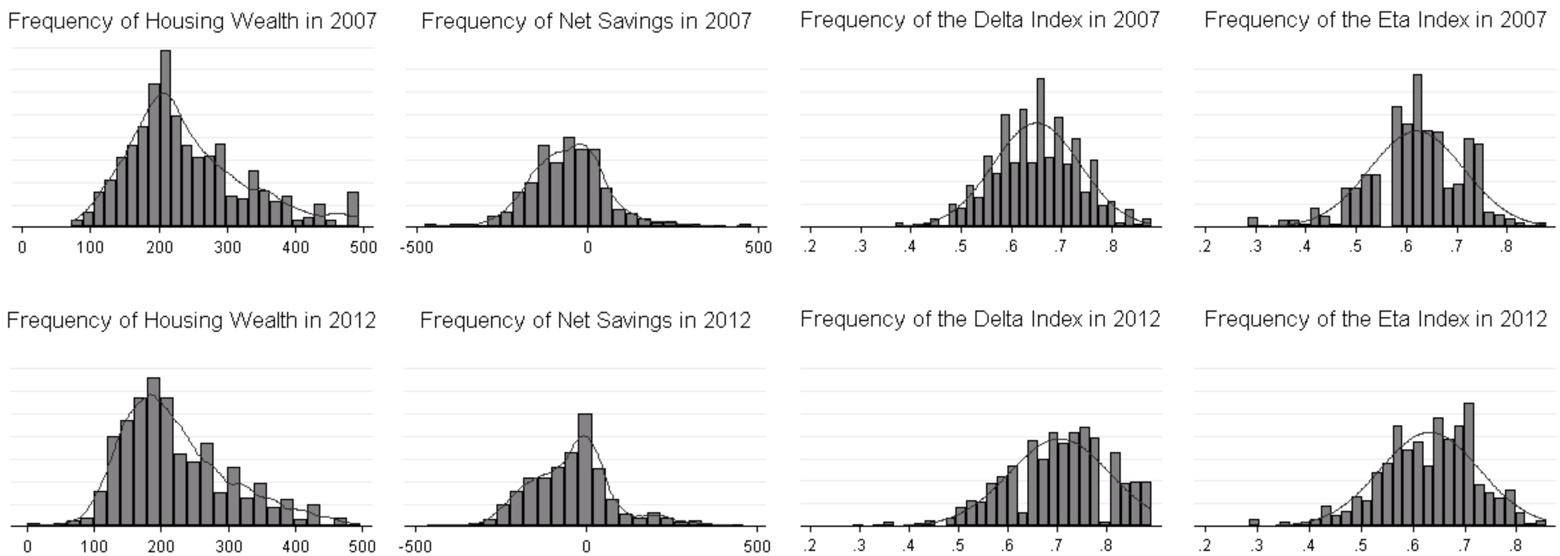


Figure D3: Comparative statics for the full model

Figure D4: Histograms of the regressors at 2007 (top) and 2012 (bottom).



Note: The normal distribution is drawn in the graphs of the indices instead of a fitted distribution. All values are in 2005 euros ($\times$1000).

Table D1: Pooled data summary.

Variable	Count	Mean	St Dev	Min	Max
Net Financial Wealth	8735	214.65	238.03	-683.22	5103.29
Net Savings	8397	-33.66	196.54	-883.22	4300.38
Housing Wealth	9038	244.91	117.70	0.87	984.61
Index δ	12267	0.65	0.11	0.14	1.00
Index η	11304	0.62	0.10	0.14	1.00
Net Income	18035	28.92	18.86	0.52	613.00
Age	18644	45.00	10.98	18	64

Note: All monetary amounts are in 2005 euros ($\times$1000). Housing wealth is self-identified. All outliers were removed as described in the text, and the composite variables were treated as missing if one of its components were missing. Values were rounded to two decimals if necessary. Net financial wealth is assets minus debt total or equivalently net savings plus housing wealth.

Table D2: Distribution of housing wealth.

Year	Mean	Percentile				
		10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
2000	209	102	141	184	256	359
2001	222	118	148	197	256	345
2002	245	116	157	210	291	420
2003	238	125	165	206	283	380
2004	240	127	167	213	274	381
2005	245	138	175	215	285	400
2006	250	142	176	221	294	392
2007	259	145	189	223	295	409
2008	257	144	184	227	302	420
2009	259	149	187	229	304	420
2010	259	143	185	231	300	416
2011	248	144	175	220	292	404
2012	233	133	165	208	277	364
Total	245	134	170	215	289	392

Table D3: Distribution of net savings.

Year	Mean	Percentile				
		10^{th}	25^{th}	50^{th}	75^{th}	90^{th}
2000	39	0	0	7	38	123
2001	-24	-148	-97	-40	6	73
2002	-39	-157	-94	-38	4	54
2003	-31	-162	-98	-36	14	79
2004	-30	-169	-109	-42	13	74
2005	-33	-180	-115	-52	10	85
2006	-32	-186	-118	-46	19	97
2007	-38	-190	-128	-53	16	85
2008	-35	-191	-131	-47	17	105
2009	-39	-197	-133	-43	18	114
2010	-40	-208	-134	-44	14	104
2011	-30	-194	-117	-25	25	119
2012	-33	-195	-124	-36	19	105
Total	-31	-181	-114	-37	16	93

Note: All monetary amounts are in 2005 euros ($\times$1000). Housing wealth is self-identified.

E Survey questions

Endogenous Variables

NET_FINANCIAL_WEALTH
- Composed from assets minus liabilities.

ASSETS
- The sum of the aggregated data on assets

  - current accounts
  - company savings accounts
  - savings and deposit accounts
  - savings account records
  - savings certificates
  - single-premium insurance
  - policies and annuity insurance
  - savings insurance and endowment insurance
  - government securities or securities accounts
  - bonds or mortgage bonds
  - stocks
  - shares bought
  - put options written or bought
  - call options written or bought
  - real estate
  - other property (cars, motor bikes, yachts, caravans)
  - money owed to family and friends
  - other possessions

LIABILITIES
- The sum of the aggregated data on debts, mortgages and other liabilities.

  - mortgages
  - personal loans
  - revolving credit
  - financing credit
  - debts with mail order companies, etcetera
  - loans with family and friends
  - student loans
  - credits on credit cards
  - other loans

NET_SAVINGS
- Composed from net financial wealth minus the expected value of the residential house (WO41).

WO41
About how much do you expect to get for your residence if you sold it today? (in thousands of euros).

Discount Rate

Answered in a scale of 1 to 7 unless noted otherwise, where 1 indicated 'totally disagree' and 7 indicates 'totally agree'. We transformed answers if necessary such that 7 implies high δ (low discounting) and 1 implies low δ (high discounting).

TOEK02
I often work on things that will only pay off in a couple of years.

TOEK03[1]
I am only concerned about the present, because I trust that things will work themselves out in the future.

TOEK04[1]
With everything I do, I am only concerned about the immediate consequences (say a period of a couple of days or weeks)

TOEK05
Whether something is convenient for me or not, to a large extent determines the decisions that I take or the actions that I undertake.

TOEK07
I am willing to sacrifice my well-being in the present to achieve certain goals in the future.

TOEK10[1]
I think there is no need to sacrifice things now for problems that lie in the future, because it will always be possible to solve these future problems later.

PLANNEN[1]
Do you find it easy or difficult to control your expenditures?
Please indicate how easy or difficult you find this on a scale from 1 to 7, where 1 means very easy and 7 means very difficult.

PERIODE1[2]
People use different time-horizons when they decide about what part of the income to spend, and what part to save.
Which of the time-horizons mentioned below is in your household most important with regard to planning expenditures and savings?
1 the next couple of months.
2 the next year.
3 the next couple of years.
4 the next 5 to 10 years.
5 more than 10 years from now.

The risk aversion index δ was constructed by taking the average of the non-missing values and dividing it by 7. If all answers were missing, the index was treated as missing.

[1] Here we reverse the answers
[2] Here we transform the answers by multiplying with 7/5

Risk Aversion

Answered in a scale of 1 to 7 unless noted otherwise, where 1 indicated 'totally disagree' and 7 indicates 'totally agree'. We transformed answers if necessary such that 7 implies high risk aversion and 1 implies low risk aversion. Only households with a net income equal or more than €10.000 had to fill in the below questions.

SPAAR1
I think it is more important to have safe investments and guaranteed returns, than to take a risk to have a chance to get the highest possible returns.

SPAAR2
I do not invest in shares, because I find this too risky.

SPAAR3[1]
If I think an investment will be profitable, I am prepared to borrow money to make this investment.

SPAAR4
I want to be certain that my investments are safe.

SPAAR5[1]
If I want to improve my financial position, I should take financial risks.

SPAAR6[1]
I am prepared to take the risk to lose money, when there is also a chance to gain money.

BESCHRYF[1] [2]
How would you describe the risks that you have taken with investments over the past few years? If you haven't made any investments, choose not applicable.
1 I have taken no risk at all.
2 I have taken small risks every now and then.
3 I have taken some risks.
4 I have sometimes taken great risks.
5 I have often taken great risks.

The risk aversion index η was constructed by taking the average of the non-missing values and dividing it by 7. If all answers were missing, the index was treated as missing.

[1] Here we reverse the answers
[2] Here we transform the answers by multiplying with 7/5

Household Characteristics

IDINK
- Net total income by households, which is composed by adding several income components such as salary and unemployment benefits. A detailed explanation can be found in the codebook online at [9])

LEEFTIJD
- Constructed from year of birth.

PARTNER
Is there a partner present in the household? (1=yes, 0=no)

AANTALKI
Number of children in the household.

GEZ5
Do you suffer from a long illness, disorder, or handicap; or do you suffer from the consequences of an accident? (1=yes, 0=no)

STED[3]
Degree of urbanization of the town/city of residence
1 very high degree of urbanization
2 high degree of urbanization
3 moderate degree of urbanization
4 low degree of urbanization
5 very low degree of urbanization

REGIO[3]
Region
1 Three largest cities
2 Other West
3 North
4 East
5 South

OPLMET[3]
Highest level of education completed
1 (Voortgezet) speciaal onderwijs / (continued) special education
2 Kleuter-, lager-of basisonderwijs / kindergarten or primary education
3 Voorbereidend middelbaar beroepsonderwijs (VMBO) / pre-vocational education
4 HAVO/VWO / pre-university education
5 MBO of het leerlingwezen / senior vocational training or training through apprentice system
6 HBO (eerste of tweede fase) / vocational colleges
7 Wetenschappelijk onderwijs WO / university education
8 Did not have education (yet)
9 other sort of education/training[4]

[3] Split among dummy variables
[4] Treated as missing