# Will we repay our debts before retirement? Or did we already, but nobody noticed?

## *The legacy of Interest-Only Mortgages, Voluntary Repayments and Saving Deposits in the Netherlands*

June 21st 2016

**Mauro Mastrogiacomo**[1]
**(DNB, VUA, Netspar)**

### Abstract

We present an analysis of the future housing debt position of (specific groups of) Dutch mortgage owners (such as starters and self-employed) around the time of their future retirement. We show that many household might be faced with an increase in housing costs, as most household debt is in interest-only loans. Their fiscal deduction for mortgages will stop 30 years after origination, in most cases this is around the mortgagor's retirement. We also show that these loans are often combined with amortizing loans in more complex mortgage structures. We acknowledge two assets that, due to data limitations of existing sources, are currently understudied: the role of voluntary repayments, and the value of the saving accounts pledged to saving/investment mortgages.

Our projections, that account for current and future non-housing wealth of mortgagors, show that individual mortgages, even if not completely redeemed, are in general not problematic for the borrowers' financial position around retirement. Housing costs will stay low if the interest-only part of debt is treated as a perpetuity, but might become a financial burden, mostly to the self-employed, who typically have less occupational pensions, otherwise. Also these debts are substantial at macro-economic level. In our most favorable simulations 1/3 of the mortgage debt existing at the beginning of 2014 will not be repaid in the next 3 decades, possibly exacerbating the banks funding-gap problem.

**JEL Classification:** C01; C23; C24; D14; G21
**Keywords:** mortgage market; interest-only mortgages; household savings; loan-to-value

---

[1] This study is based on DNB research carried out and supervised by Mauro Mastrogiacomo. Some results of this project have already appeared in the Overview Financial Stability 2014 published by DNB and in the master thesis of Jan Jakob Lameijer, who was supervised by Mauro Mastrogiacomo and Rob Alessie at DNB and Groningen University (Master's Thesis in Econometrics *s1904205*)

# 1 Introduction

About 60% (7%) of the Dutch mortgage portfolio consists of IO (investment) loans (Mastrogiacomo and van der Molen, 2015), these were popular because they allowed low monthly payments and because of the generous mortgage interest deduction (MID). Policy intervention in 2001 tried to limit their development by first stopping the MID after 30 years, and in 2013 by making IO loans no longer eligible for these fiscal rebates. Especially in the long run, the large amount of debt in this type of loans could impose a financial burden to households, when the maximum amount of years that owners are entitled to the MID will be exceeded. Among the future unknowns stands the possibility to treat IO loans as IO perpetuities (low financial burden for households) or to transform them in annuities with short maturity (higher financial burden). For households, this could imply an increase of their debt service to income ratio (DSTI) as both the net monthly costs could increase, and because around that time many will be about to retire (thus also facing an income reduction). For specific groups, such as the self-employed, this might become problematic as they are observed having larger debt and their future pension annuity should be low, as they did often not provide for a private pension (Mastrogiacomo et al 2014). Also a substantial amount of debt could be left in the banks' books beyond maturity.

Are we going to pay back our debt before retirement, given the high IO share in our mortgage? And if we do not, how would mortgage costs increase if outstanding debt at maturity is then treated either as an IO perpetuity or as an annuity?

We can answer these questions using the loan level data (LLD) gathered by DNB, as we are able to disaggregate housing wealth, quantifying the accumulated savings and assets pledged as collateral for the mortgage in all periods preceding maturity. The advantage of this novel dataset is that we observe detailed information on individual loan characteristics, where households typically have multiple loans to finance the house. So, more specifically, we aim to find an answer to the following question: *what are the risks in the long run associated with the large share of interest-only (and investment) loans in the Dutch mortgage portfolio and how do these risk differ across specific groups in the population, specifically the self-employed*? In order to provide a more specific answer to the rather broad question above, we formulate three sub-questions: 1.How are IO loans distributed across the Dutch homeowners? 2.How much of the current mortgage debt will be redeemed in the coming thirty years? 3.Will different types of households with IO loans have saved enough to pay off their mortgage at maturity? If not, how large will housing costs be after maturity or retirement?

These questions are broad in scope but at the same time also delimit our research. We are not enquiring the optimal allocation of debt over the life-cycle. Such a discussion involves the consideration of the optimal path of retirement savings, including housing savings (Sun et al, 2007) as well as their fiscal treatment (Bovenberg and Jacobs, 2008), which in the Netherlands could be achieved by making more often use of reverse mortgages (Dillingh et al, 2015). So, we do not suggest that one should repay back the whole mortgage, but we want to understand how large outstanding debt will be after mortgage maturity and around retirement.

To answer the questions above, we build a microsimulation model that simulates the mortgage debt at borrower level up to thirty years in the future, where we use 2014 as the starting year and 2043 as the last year. We estimate a model for voluntary repayments and show that they contribute substantially to the redemption of the current mortgage debt. Furthermore, the contractual mortgage repayments and capital accumulation on accounts pledged to the mortgage are modeled deterministically, based on some quite undisputed assumptions. This shows how part of the outstanding debt was already repaid, while another part is very likely to be repaid soon.

In order to show the heterogeneity in our population, we will highlight two specific dimensions. First, we separate different cohorts of borrowers. Second we will isolate those who were self-employed at mortgage origination, as they are more likely to have low contributions in the second pillar. Also, we will discuss the amortization of investment loans.

We find that most interest-only loans are combined with amortizing loans, but where still 36% of the borrowers have a full interest-only mortgage. However, these are mostly older borrowers having substantial home equity. Starters are hardly represented in the 100% IO category. When we weight this share by household debt, we do not find substantial differences between wage-employed and self-employed. Mortgages that are currently underwater are typically amortizing mortgages (at least partially). We find that the share of underwater mortgages will decrease even if house prices stay constant for the coming thirty years. Only when house prices decrease with more than 2% annually and no voluntary repayments are made, we find that both the share of underwater mortgages and the average LTV will increase. Problematic groups are the self-employed and the owners of investment loans. We observe for the self-employed a significantly higher LTV ratio (that will drop below 60% about 10 years later, relative to the whole population). We also find that investment-loans-owners have chosen to complete their mortgage combining the investment loan with an IO loan. They repay thus very little of their mortgage, and the performance of their investments has in the last decade been below the one projected upon signing the loan. At the same time, we find that almost all mortgages will be above water at maturity and that most mortgages with high LTV ratios are backed by the government with a national guarantee (though the guaranteed amount diminishes over time following an annuity scheme).

Another contribution of this study is that it relates mortgage debt to non-housing wealth. Using a second administrative dataset we estimate a model for wealth based on a subset of covariates that are observed in the LLD as well. We find that many borrowers with a remaining debt at maturity will not have saved enough to fully repay their mortgage at maturity. Especially the mortgages originated around the bursting of the housing bubble will have a remaining debt of roughly 90 000 euro on average, and financial wealth of about 30 000 euro. This figure is heterogeneous across the population. For instance, self-employed will have a remaining debt of about 150 000 – 200 000 euro, though larger financial assets (60 000 euro). Moreover, as retirement is likely to occur soon afterwards, these borrowers may be confronted with a drop in income as well. This drop may be more severe for self-employed workers, excluded from second pillar savings, that did not prepare for their retirement by making additional savings.

It is thus unclear what the future housing costs will be of borrowers that do not repay in full at maturity, as it is not automatic that the remaining IO debt will always be treated as an IO perpetuity. If it is not, the most exposed households may end up facing average increases of their monthly payments by amounts larger than the current social security benefit. Here, we focus on the period around maturity as that is often close to retirement. However, the debt position of households is relevant also before. Being underwater is a risk trigger that gets activated in association with several shocks that one can possibly envisage. Unemployment, bankruptcy, divorce and disability are the most relevant from an individual perspective, while an interest rate shock could affect all borrowers. When these shocks materialize, borrowers with underwater mortgages become more financially distressed. If their number is large (as it was in 2013 when 1/3 of mortgages was underwater in the Netherlands), the combination of these risks may result in additional defaults or in economic downturns. The first did not happen during the last crisis, but consumption dropped considerably (Verbruggen et al, 2015), which was one of the main causes of the recent recession (Mian and Sufi, 2015, have quantified this for the US).

The remainder of this study is organized as follows. The next section discusses the most important features of the Dutch mortgage market; after which we provide a description of the datasets in Section 3. The econometric models and estimation procedures are presented in Section 4, together with an overview of the design of the microsimulation model. Next, we present and interpret both the estimation and simulation results in Section 5. Section 6 concludes this study.

## 2 Characteristics of the Dutch mortgage market

The Dutch housing market has undergone dramatic changes over the last two decades. An unprecedented growth in house prices in the latter half of the 1990s was associated with rising household leverage. This became possible when banks, supported by policy makers and public opinion, started to take the income of the partner into account when assessing the borrowing possibilities of households, thereby relaxing credit constraints. Banks allowed borrowers to increase their mortgages due to the expected increase in collateral value and, in turn, households used their extended borrowing capacity to accumulate debt mainly for housing purposes.[2] The higher demand for housing and loosening of credit constraints, along with the inelastic supply, caused the house prices to increase even further. This procyclical phenomenon, referred to as the collateral amplification mechanism or, more in general, the financial accelerator (Almeida et al. (2006), Bernanke et al. (1996)) has been the root cause of credit crises all around the world (for further reading, see for instance Kiyotaki and Morre (1997), Lorenzoni (2008)). Especially in the Netherlands, where the mortgage interest payments are fully tax-deductible, households were encouraged to finance their house with debt. The introduction of the National Mortgage Guarantee (NHG [Nationale Hypotheek Garantie]) in 2000, where government acts as guarantor for the mortgage payments, allowed banks to ease the credit constraints for households even further. NHG can only be issued to mortgages for houses with transactions prices below a legislated threshold.

Eventually, the bursting of the housing bubble in 2008 revealed the vulnerabilities of the Dutch housing market. By 2013, house prices had decreased by more than 20% compared to the peak in August 2008. During the same period the number of Dutch mortgages that were underwater increased from 10% to approximately 30% (DNB, 2014).

Both the decrease in house prices and increase in mortgage debt have contributed to a higher loss given default (LGD), resulting in substantial credit risk for banks. A forced sale after the crisis is no longer enough to cover the outstanding mortgage debt (on average, the foreclosure value in the Netherlands is approximately 85% of the market value). Moreover, banks have become highly dependent on (short term) market funding due to the shortage of savings deposits as a stable funding source, resulting in a large deposit funding gap (DFG). This maturity mismatch between assets and liabilities becomes in particular troublesome when markets are not performing well, such that refunding will be harder. One way to overcome this problem is to securitize part of the mortgage portfolio via the residential mortgage-backed securities (RMBS). Unfortunately, this type of funding has become much more expensive because investors

[2] In 2000, mortgage interest deductibility was restricted to buying or renovating a house, encouraging households to use the credit mainly for housing and home improvements.

have become aware of the risky mortgage portfolio (Jansen et al., 2013). In effect, the European Union is now considering tightening the eligibility rules into the RMBS pool, by for instance only allowing mortgages with an LTV below a conservative threshold (say 80%). Finally, part of the credit risks faced by banks are transmitted to the government via the NHG.

In reaction, new regulations were implemented to reduce these risks and to prevent excessive credit growth. In 2013, the Dutch government introduced the rule that only new fully amortizing mortgages are eligible for the interest deduction. The maximum tax-deductibility will be gradually reduced from 52% in 2014 to 38% in 2042, which also applies for existing mortgages. Furthermore, an upper limit to the LTV for home buyers was initiated. In 2015, this LTV cap was set to 103%, which will gradually reduce to 100% in 2018. Also, the Financial Stability Committee (FSC) has adviced lowering the limit even further to 90%. One last regulation to keep in mind is that from October 2013 until December 2014 the government temporarily raised the exemption from gift taxes to 100 000 euro, but only when the money is used for mortgage redemption or home-improvements. At the same time, most lending institutions also increased the maximum amount that can be voluntarily repaid without incurring a penalty. This means that the Dutch government has chosen to use a strong fiscal stimulus to induce new borrowers into buying annuities rather than IO loans. Policy options for current mortgage owners, such as nudging them into choices that the government finds optimal, have not yet been attempted. With the recently falling interest rates for instance, upon resetting mortgage contracts, retirement saving programs like Save More Tomorrow could have been used on the mortgage market. Think for instance of directing the gains of lower interest rates in the direction of higher repayments.

# 3 Data

## 3.1 Loan Level Data (LLD)

The LLD is collected by DNB using the reporting template for Residential Mortgage-Backed Securities (RMBS) of the European Data Warehouse.[3] In order to use a securitized mortgage as collateral, each lending institution must agree to the 100% transparency policy of the ECB and fill in the template. The DNB version of the LLD also includes the back-books on top of the securitized pool discussed above, which the institutions deliver on voluntary basis. This is essential, as securitized mortgages in the Netherlands are not a random sample of the mortgage portfolio, and are typically rated AAA. Although the LLD meets the reporting requirements

[3] The RMBS template can be found at https://www.ecb.europa.eu/paym/coll/loanlevel/transmission/html/index.en.html (accessed on 11-01-2014)

of the ECB, it is to some extent not designed for analytical purposes. Mastrogiacomo and van der Molen (2015) describe some limitations and advantages of the LLD.

[Table 1 here]

The first wave was collected in 2012 Q4 and the last currently available wave is 2013 Q4. Table 1 testifies of the main advantages of the LLD. First, from the total number of borrowers and loans reported in the table we see that a mortgage typically consists of multiple loans (approximately two loans per mortgage on average). Observing each loan and borrower separately allows, for example, to accurately determine the repayment schemes of each loan, the debt-weighted share of interest-only mortgages and to impute the saving deposits pledged to each loan. The table shows that roughly 60% of the loans are IO, in accordance with the aggregate figures reported in the literature. Due to the granularity of our data we can nuance this large portion of interest-only loans. As shown in the table, we observe that only 35% of the borrowers have a full interest-only mortgage, meaning the remaining borrowers amortize at least to some extent. In the next section we will also present the debt-weighted shares per loan type, which provides a more complete picture.

We estimate that the LLD covers approximately 80% of the total population, as shown in Table 1. For each loan record in the LLD a large number of attributes is reported. Each record includes a unique loan and borrower identifier, which allows tracking them over time if (and only if) the borrowers stay within the same bank.

Further, some banks apparently observe the assets pledged to the mortgage and subtract this from the outstanding debt. This is different from monetary statistics practices, where the two accounts are kept separately. It is not immediate to distinguish between voluntary and contractual repayments when amortizing loans are present. In order to break this observational equivalence, we make use of the panel nature of the data. By looking at the difference in loan balance over all five waves, we are able to identify the flow into the accumulated capital (AC) pledged to the mortgage.[4] This means that we are dealing with two definitions of mortgage debt at the same time. A gross definition, where the AC is not considered and a net definition that subtracts the AC. Fortunately, the large number of attributes in the LLD allows to estimate the AC for each loan, such that we are able to approximate both gross and net mortgage debt.

Several value concepts could be used to determine the value of the property, such that care must be taken when comparing LTV ratios in the literature. The fair market value might differ

[4] Specifically, the flow is identified by the regularity in the decreases of the outstanding debt. Everything on top of this qualifies as a voluntary repayment.

from the actual transaction price due to market distress and inefficiencies. Other commonly used value concepts that differ from the fair market value are, for example, the tax assessed value (WOZ-value [Waardering Onroerende Zaken])[5] determined by the taxing authority and the liquidation value.[6]

[table 2 here]

The LLD does not necessarily allow for a consistent definition of the valuation amount, as different value measures are used across observations. From Table 2 we observe that for more than 50% of the properties the appraised value is reported, where the appraisal is performed by an expert. The purpose of the appraisal, however, is not indicated, but perhaps we can learn more by comparing the average property values resulting from the different valuation methods. As can be seen, the average property value determined by an expert inspection is somewhat smaller compared to the WOZ-value and the value determined by an estate agent. This might indicate that experts indeed valuing the property as collateral for the mortgage, where the sale needs to be achieved quickly, leading to a more conservative valuation. However, here we make the assumption that the valuation method is chosen randomly, which does not have to be the case.

Mortgage debt concepts are also slightly different in both datasets. First, the IPO reports only a gross definition. The approximated gross mortgage debt in the LLD is possibly an underestimation. Second, the IPO only reports the fiscal debt, which is the part of the mortgage debt used to finance the prime residence and for which the interest payments can be deducted from taxable income. In our LTV definition we will use the net mortgage debt, as it provides a more complete picture of the financial position and risks of the households.

### 3.1.1 Descriptive statistics

This subsection presents some descriptive statistics based on the 2013 Q4 wave. After removing borrowers with missing or highly unrealistic values for the relevant variables, we are left with 2 375 545 borrowers having 4 521 284 loans in total (for 472 991 of the removed borrowers the birth year was missing). Using this restricted sample we estimate the aggregate gross mortgage debt in the Netherlands to be approximately 639 billion euro. Subtracting the estimated 30 billion euro AC (which is possibly an underestimation, as will be discussed below)

[5] Historically, the WOZ-value was an underestimation of transaction prices, whereas the two have become more aligned in more recent years.

[6] In the Netherlands, a foreclosure auction results on average in a liquidation value of 80% of the market value.

yields an estimate of the net mortgage debt of 609 billion euro.

The pie chart on the left-top in Figure 1 presents the debt-weighted share of each mortgage type. Similar to Table 1, we find that almost 60% of the net mortgage debt comes from interest-only loans[7]. The bottom of the figure is dedicated to those self-employed at origination. It shows that there the IO shares and mortgage types of this sub-group do not differ from those of the rest of the population.

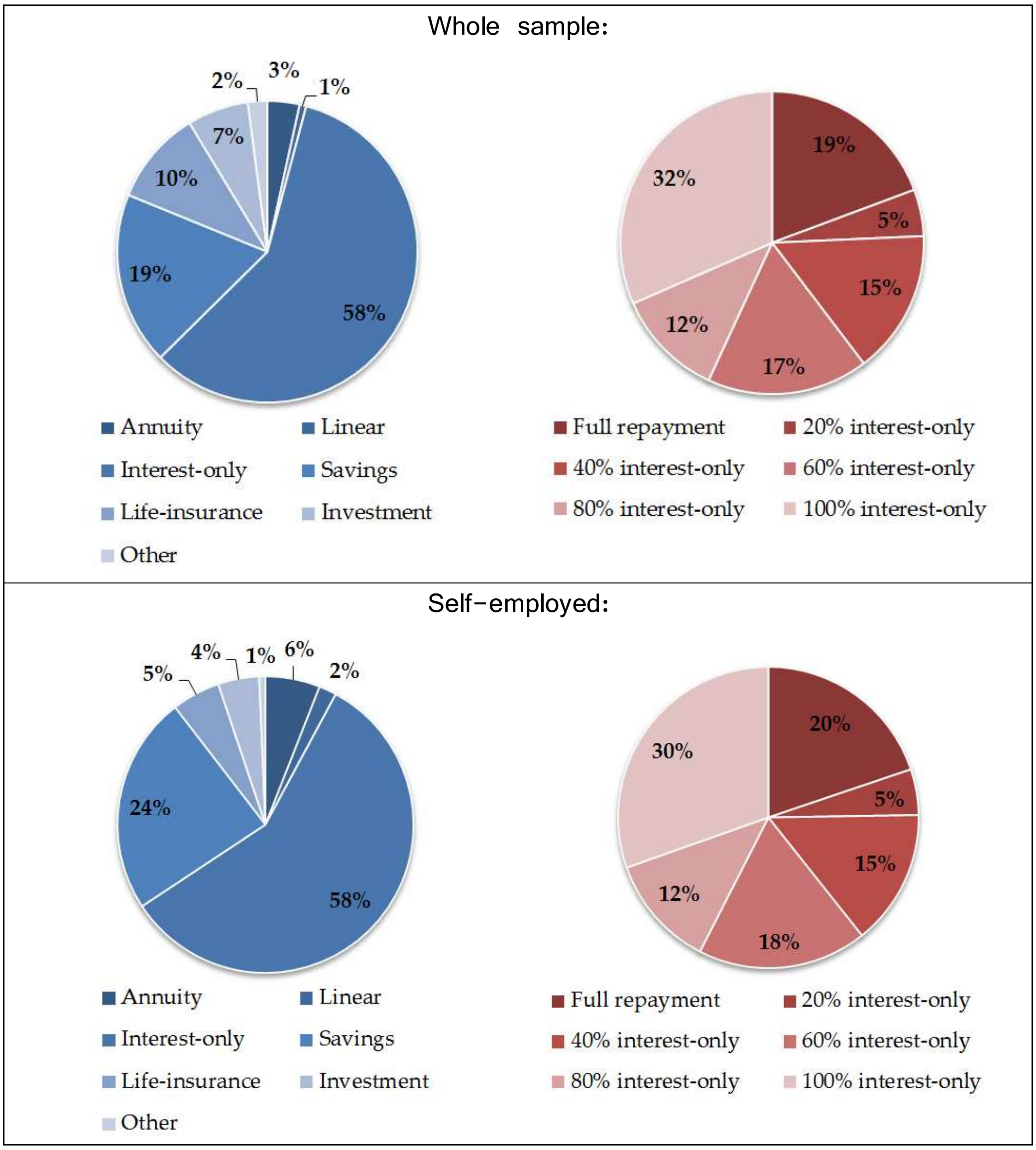



**Figure 1:** Debt-weighted share of each mortgage type (left) and the debt-weighted share of borrowers per interest-only category (right), both based on net mortgage debt in 2013Q4.

[7] The difference between the 50% indicated in Table 1 can be attributed to the difference in net and gross mortgage debt (CPB, 2014).

From Figure 1, we observe that 32% of the net mortgage debt comes from borrowers having a 100% interest-only mortgage, which is even less than the 35% from Table 1. This result therefore shows that the large part of interest-only mortgages are often combined with other amortizing mortgages. Descriptive statistics for the relevant variables are presented in Table 3, where descriptives are given per interest-only category.[8] The statistics are given on borrower-level, where the interest rate is the average debt-weighted interest rate of all mortgage loans of the borrower.

We observe that the relationship between interest-only share and LTV is not linear. On average, the youngest borrowers fall in the 40% IO category, where the average LTV is no less than 93% and where 54% of the mortgages are underwater. These borrowers do contractually amortize on more than half of their mortgage debt. Also, we find that a large share of underwater mortgages is often accompanied with a large share of mortgages that are NHG-guaranteed.

[table 3 here]

Voluntary repayments are not directly observed, but can be retrieved by taking the difference in mortgage balance between the beginnig of 2013 and of 2014, where we correct for contractual mortgage repayments.

By taking the yearly difference we remove all seasonal components. However, given the limited number of waves in the LLD we can only calculate a proxy of the voluntary repayments for one specific year. We should keep in mind that for the last two months of that period (starting from October 2013) the exemption from gift taxes was raised to 100 000 euro for home-related expenditures.

Considering the administrative costs of processing repayments, most lending institutions have set a lower limit. Therefore we treat all voluntary repayments (calculated drops in outstandig debt above amortization) of less than 2 000 euro as zero, as we wish to capture the true underlying distribution (which we only observe for voluntary repayments above 2 000).

[8] The table does not contain the variable income, which might be considered a relevant variable as it probably has a strong effect on both savings and voluntary repayments. Unfortunately, income is reported for only 50% of the borrowers, and a comparison of means test strongly rejects the hypothesis that these observations are missing at random.

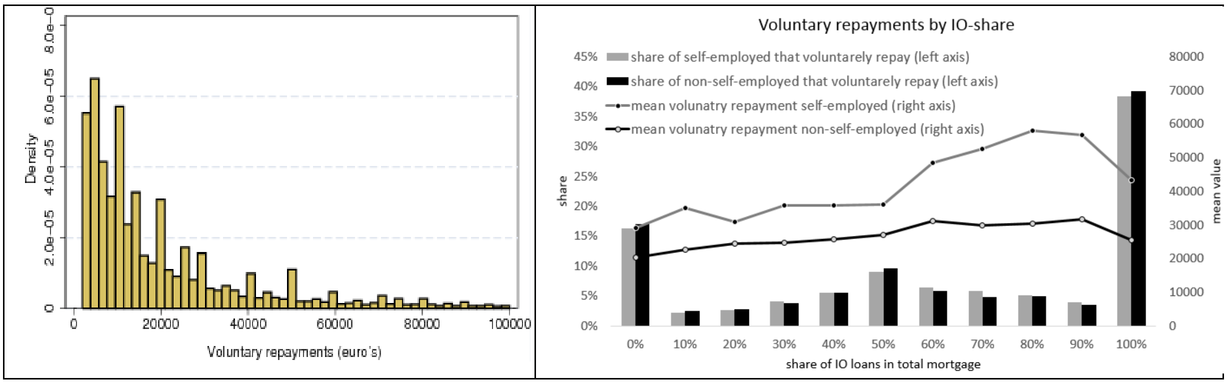



**Figure 2:** Distribution of the voluntary repayments in 2013 (truncated at 100 000 euro), and voluntary repayments by IO share.

As a result, we find that 13.74% of the borrowers in our sample have made a voluntarily repayment on their mortgage in 2013. The sum of these repayments is estimated to be 13.36 billion euro at aggregate level, representing roughly 2% of the net mortgage debt. A histogram of the resulting (non-zero) voluntary repayments is provided in the left panel of Figure 2. The peaks indicate that round numbers are more popular amounts to voluntarily repay, as expected. The right panel of the figure shows that voluntary repayments are not unifrom across the population. As an illustration we show that the share of those repaying in 2013 differs dependning on the IO-share in the mortgage, with fully IO mortgages repaying more often. Also we break down downpayments to occupation. The share of repayments is not different when we look at self-employed or at non self-employed, but the mean repayment is higher for self-employed (who have larger debt) and increases with the IO share.

## 3.2 Income Panel Study (IPO)

To analyze non-housing wealth we use seven waves of the IPO dataset (2005 - 2011) gathered by the CBS. In total, the dataset consists of 1 852 323 observations, containing information on 112 942 unique households. We only select the household heads that own a property financed by a mortgage. We estimate the mortgage interest rate by dividing the yearly mortgage interest payment by the gross mortgage debt. Subsequently, we remove observations for which the resulting interest rate is unrealistic (less than 1% or exceeding 10%). The selected sample consists of 341 118 observations on 63 791 unique borrowers.

[table 4 here]

The missing information in the LLD that is provided by the IPO dataset is non-housing wealth. Specifically, we are interested in the net household savings, which we define to be the sum of all non-housing financial assets (savings and investment accounts not pledged to the mortgage, where shareholdings with substantial business interest are not considered) minus all outstanding debt balances other than the mortgage debt. Unfortunately, the LLD and IPO do not contain unique borrower identifiers by which the datasets could be matched. We aim to estimate a model for net savings based on variables that are observed in both datasets and use the resulting model to estimate net savings in the LLD. Descriptive statistics of all common variables and net savings are presented in Table 4, for three of the seven waves. Especially the large standard deviation and relatively large difference between the mean and median of net savings are notable. As will be discussed later, they alert us that difficulties may arise when modeling net savings.

Figure 3 presents age and cohort patterns of the net savings, where we use five-year birth cohorts.

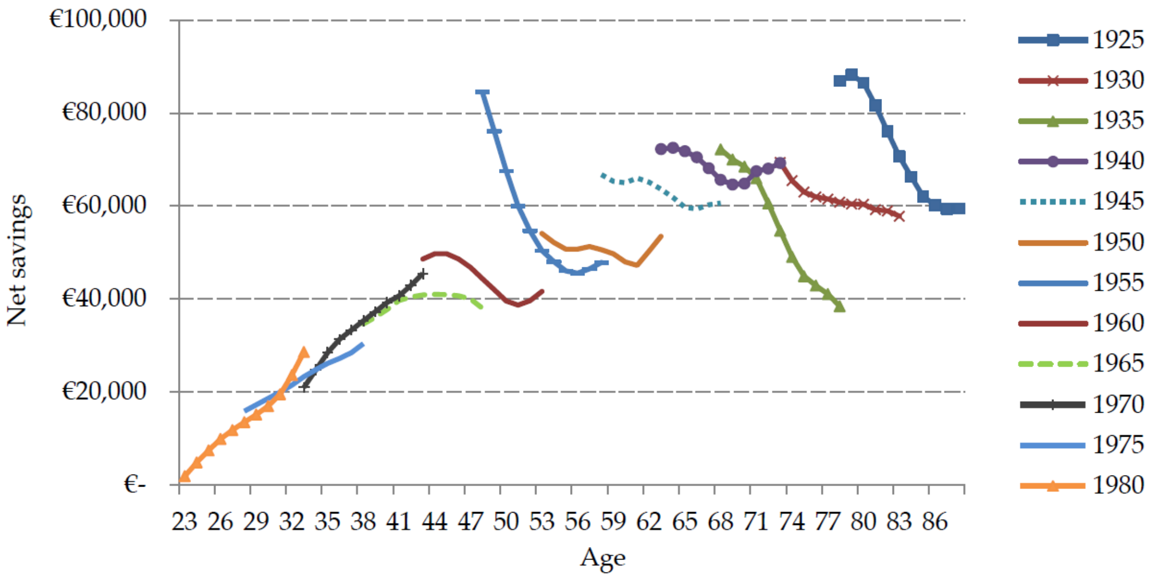



**Figure 3:** Net savings by birth cohort using nonparametric locally weighted regression (LOWESS) with a bandwidth of 0.8. Labels correspond to the middle year of each cohort.

Birth years 1923 - 1927 are for the oldest cohort and 1988 - 1992 for the youngest cohort, where the labels correspond to the middle year of each cohort. To enhance visual information we have fit a LOWESS curve (Cleveland, 1979) with a smoothing parameter of 0.8 for each cohort. We observe an increase in net savings over age for young cohorts and a decrease for older cohorts. Differences in average net savings between cohorts at the same age are indicated by vertical differences between the cohort curves.

Figure 4 compares the distribution of the property value as observed in both the 2011 wave of the IPO and the 2013Q4 wave of the LLD, where the values are indexed to 2011 for the latter dataset. Reassuringly, the distributions are very similar. Also, a comparison of the distribution of the gross mortgage debt is provided in Figure 5.

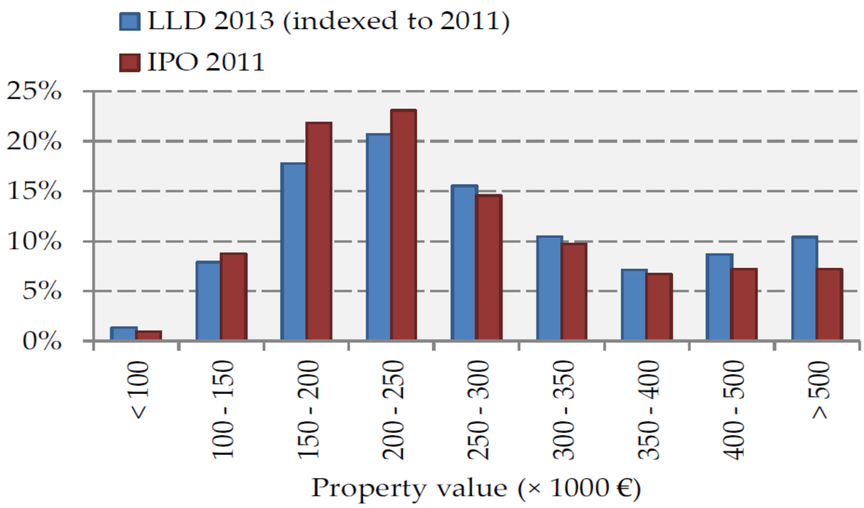



**Figure 4:** Distribution of the property value in both LLD and IPO

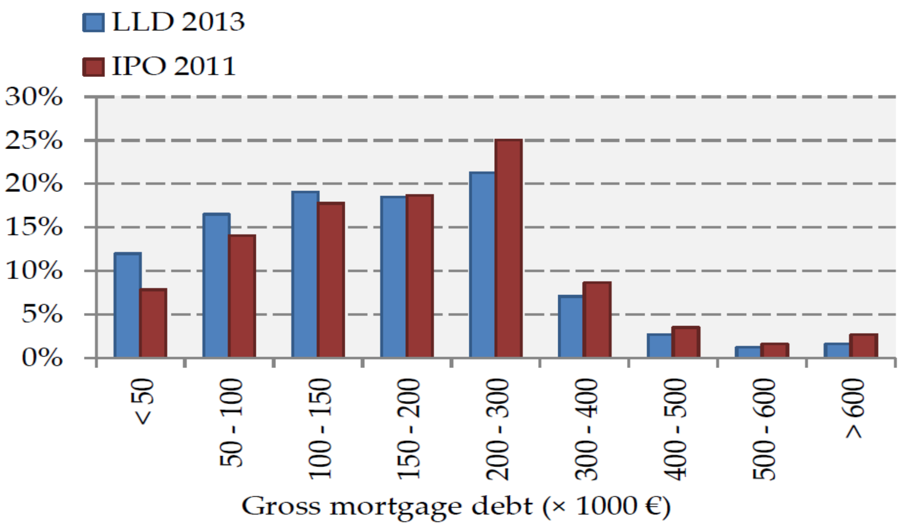



**Figure 5:** Distribution of the gross mortgage debt in both LLD and IPO

Finally, both the LLD and IPO report the first two numbers of the postal code. This allows us to impute some variables based on postcode-level in both the IPO and LLD, such as the debt-weighted share of interest-only mortgage per postcode, the average property value per postcode and the number of real estate transactions per postcode. The former two are obtained from the LLD and the latter from the Dutch Association of Real Estate Brokers and Real Estate Experts (NVM [Nederlandse Vereniging van Makelaars o.g. en Vastgoeddeskundigen]).

# 4 Methodology

## 4.1 Microsimulation model of the Dutch mortgage portfolio

We differentiate between three components that jointly determine the net mortgage debt of each borrower: 1) the periodic mortgage repayments as contractually specified, 2) the capital accumulation on accounts pledged to the mortgage, and 3) voluntary repayments. The first two components are modeled deterministically, based on the structure of the mortgage and some assumptions, whereas the latter component is modeled stochastically. We focus therefore in the Appendix on the latter, while here we describe the simulation method only.

We start our simulation in 2014 using the borrowers from the LLD observed on December 31st 2013. To alleviate computational intensity we select a random subsample of 50 000 borrowers. For these borrowers we simulate the mortgage debt and net savings for the upcoming thirty years, where 2043 is the last simulated year. A general overview of the simulation procedure per borrower is provided in Figure 6.

The first step in the microsimulation is to simulate the voluntary repayments for the upcoming year (2014). Anticipating on the estimation results, this will be done according the Cragg log-normal hurdle presented in equation (6) in the Appendix. First, to simulate the participation decision, we draw a random value from the uniform distribution for each borrower. Only if this random variable is less than the predicted value from the probit model (Part I), the borrower voluntarily repays. Next, to simulate the amount of the voluntary repayment we use the predicted value from Part II of the log-normal hurdle, where repayment shocks are drawn from the normal distribution with zero mean and variance $\hat{\sigma}^2$. Here, $\hat{\sigma}^2$ is the estimated variance of $u_{it}$ from equation (6). Finally, the exponential function is used to transform the repayment amount back to levels.

Now that we have simulated the voluntary repayments in 2014, we can update all other debt-related variables (total net debt, debt-weighted share of interest-only loans, LTV, etc.). Here, we assume the voluntary repayments are first used to repay the interest-only loans. If the borrower no longer has interest-only loans, the repayments will be used to repay mortgage loans for which capital is accumulated in a separate account (investment[9]/savings/life insurance). The voluntary repayments will only be used to repay amortizing mortgages (annuity/linear) in case the borrower has no other mortgage loans. The contractual mortgage repayment and capital accumulation are calculated as described above. Furthermore, we make a few assumptions on the change in property value, GDP and CPI. The basis scenario assumes constant house prices and a yearly 2% increase in both GDP and CPI. To test the sensitivity of the results to these assumption we experiment with yearly house price changes of 3% and -2% and with GDP and CPI changes of 4% and -2%.

Recursively estimating the voluntary repayments and updating the values of the variables until 2043 completes the simulation.

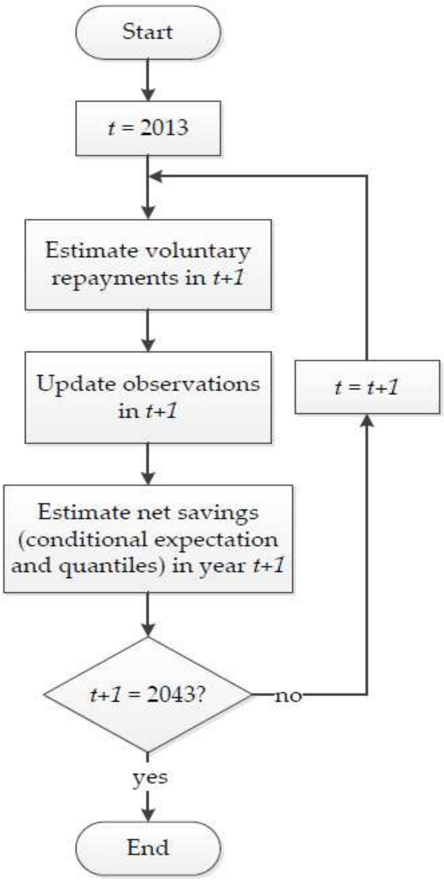



**Figure 6:** Flowchart of the microsimulation per borrower

[9] Notice that it is typically not allow to repay an investment loan without penalties, as this would be equivalent to withdraw money from the investment found.

[table 5 here]

[table 6 here]

[table 7 here]

# 5 Results

## 5.1 Estimation results

### 5.1.1 Voluntary repayments

The first two columns in Table 5 show the estimated coefficients and associated ME of the probit model on the decision to voluntarily repay. Partly due to the large sample size, all coefficients and ME are statistically significant at a 1% level. Unfortunately, as indicated by the low value of the pseudo $R^2$, the model fits rather poorly. Much of the variance in the choice to voluntarily repay is still not explained by the regressors, this implies that in the simulation much of the results will be driven by the random draws of the unexplained part. Also, for the same reason, we do not discuss the economic interpretation of the results, and use the model only as the scoring device needed in the simulation method The LM tests strongly reject the hypothesis of homoskedasticity and normally distributed error terms. Again, the rejection might be attributed to the large sample size. To investigate the scale of this problem we compare the ME of the probit model with ME resulting from a linear probability and logit regression. If the assumptions on the error term are wrong, the ME should differ substantially as the underlying distributional assumptions differ across the models. The estimation results for the three probability models are presented in Table 6. The estimated ME are very similar for all three models, thereby providing an incentive to assume that the probit model is correctly specified, although the heteroskedasticity and normality tests are rejected.

The third and fourth column in Table 5 present the estimation results of the Tobit model, where voluntary repayments in levels is the dependent variable. If the Tobit model is correctly specified, the probit and Tobit model should yield similar estimates of the ME. However, we observe that the ME of interest rate and current LTV are different in both sign and magnitude. When transforming the data using the natural logarithm, we find that the distribution is almost symmetrical (skewness=0.37) with negligible non-normal kurtosis of 2.75. The estimated Tobit model of voluntary repayments in logs is provided in Table 5 as well. As can be seen, the estimated ME are now much more similar to those of the probit model (all estimates have the same sign, but the magnitude of the ME of interest rate and current LTV is still different). Also, the Tobit model in logs fits the data considerably better in terms of both pseudo $R^2$ and log-likelihood (although the $R^2$ is still very low).

Finally, the last column of reg-level presents the estimation results of Part II of the Cragg log-normal hurdle. We find that the estimated Cragg log-normal hurdle yields the same pseudo $R^2$ as the Tobit in logs, but has a larger log-likelihood. We thus choose to model the voluntary repayments using the Cragg log-normal hurdle. Additional, to allow for variation in coefficients between mortgages with different shares of interest-only loans, we fit a Cragg log-normal hurdle for all six interest-only categories as defined in subsub:descriptives separately. By doing so, we also allow the variance of the error terms in both parts of the Cragg log-normal hurdle to be different for all interest-only categories (i.e. we partly allow for heteroskedasticity).

### 5.1.2 Net household savings

The first column in Table 7 presents the estimation results of the robust regression on net savings, where panel-robust bootstrap standard errors are used. Not all variables are statistically significant, but we choose not to exclude any of the regressors from the model.

We want to use every variable that the IPO and LLD have in common to estimate net savings in the LLD. Remarkably, the birth cohorts (and postcode variables) are jointly insignificant, which contradicts the visual information from Figure 3.

The estimation results of the three quantile regressions are provided in Table 7 as well, where the dependent variable is the IHS transformation of net savings. Recall that we do not present bootstrap errors for this regression due to the computational intensity. As a result, the cohort and postcode variables incorrectly appear jointly significant (a robust regression on net savings without bootstrap standard errors yields jointly significant cohort effects at a 1% significance level as well). In spite of the large number of regressors included in the model, the quantile regressions still fit rather poorly as indicated by the low $R^2$ values.

## 5.2 Simulation results

First, we focus on how much of the current mortgage debt will be redeemed in the coming thirty years. Figure 7 presents the simulation results of the aggregate net mortgage debt for different scenarios. The upper line represents the scenario where borrowers do not make voluntary repayments, which provides a quick check on whether we have modeled the contractual repayments correctly. In this scenario, roughly 33% of the current mortgage debt will be redeemed in 2043. Indeed, the remaining 67% comes from all interest-only loans (58%), investment loans (7%) and loans classified as "other" (2%), for which we assumed no capital is accumulated for the moment. Later we relax this assumption. If we treat the latter two types similar to savings mortgages, we find that 42% will be redeemed in 2043 rather than 33% (not presented in the figure).

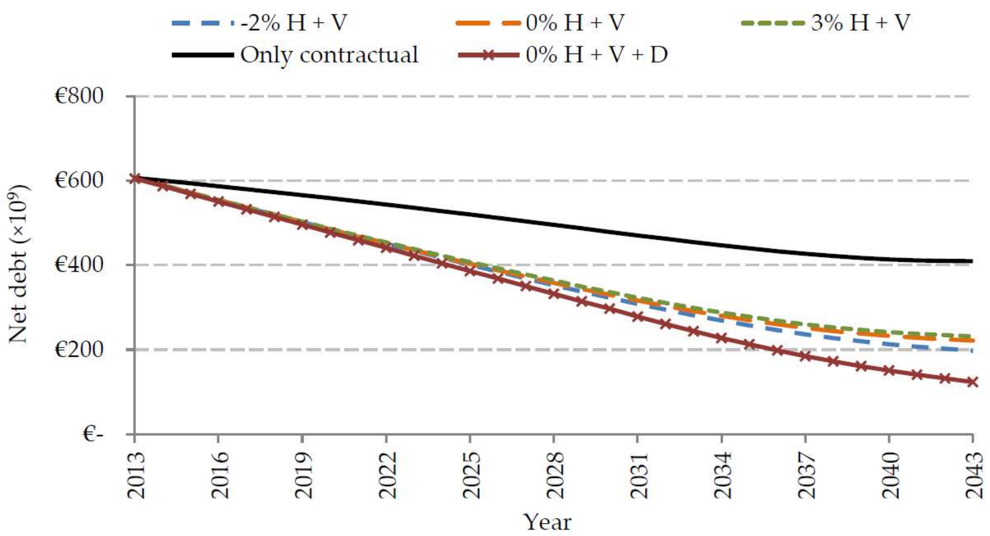
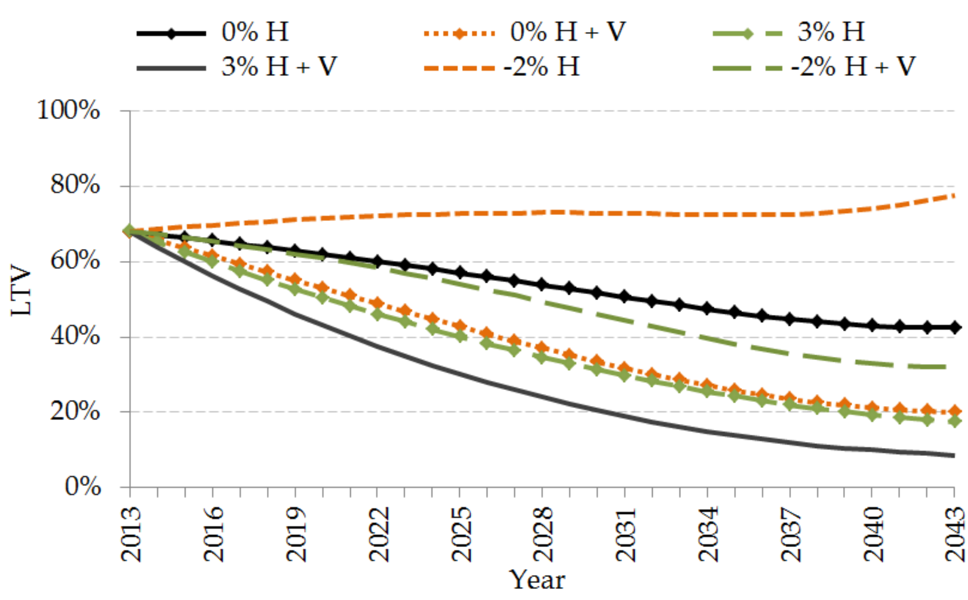



**Figure 7:** Simulation of the aggregate net mortgage debt for currently existing mortgages in the Netherlands. Different scenarios are considered (H = house price change; V = voluntary repayments; D = mortgage is repaid at death (85 year)).

Whole sample:




Self-employed:



**Figure 8:** Simulation of the average LTV of the mortgages currently existing in the Netherlands, where different scenarios are considered (H = house price change; V = voluntary repayments).

The three dashed lines in Figure 7 allow for voluntary repayments, where different house price scenarios are considered. We observe that voluntary repayments contribute substantially to the redemption of the mortgage debt; almost half of the redeemed mortgage debt in 2043 comes from voluntary repayments. As can be seen, this result is not very sensitive to different house price assumptions. Additionally, the marked line shows that another hundred billion euro will be redeemed when taking mortality into account. But older borrowers typically have substantial home equity. Consequently, only 0.7% of these borrowers are underwater when reaching the age of 85, where we assume constant house prices. Hence, the losses incurred by the lending institutions are probably very limited. Nonetheless it is likely that about 1/3 of currently outstanding debt, will not be repaid in the coming 30 years, thus aggravating the funding gap problem of banks, described above. The development of the average LTV is presented in Figure 8, where different house price scenarios are considered. We also highlight at the bottom of the figure that self-employed start with a higher LTV ratio. It takes this group longer to reduce the LTV below given thresholds (almost 10 years longer to reduce it below 60% for instance) but thanks to their repayment behavior they also will finally end up with similar indebtedness as the whole sample in most scenarios.

Figure 9 shows the evolution of the share of underwater mortgages. Mortgages currently underwater are typically amortizing mortgages (at least partially), such that the share of underwater mortgages will decrease even when considering constant house prices. In the most optimistic scenario we find that almost all mortgages currently existing will be above water in 2022. In that same year, only 6% will be underwater when house prices remain constant and voluntary repayments are allowed. Only if house prices decrease with 2% annually and voluntary repayments are not considered we observe an increase in both average LTV and underwater share. Both figures again show that the contribution of voluntary repayments is substantial.

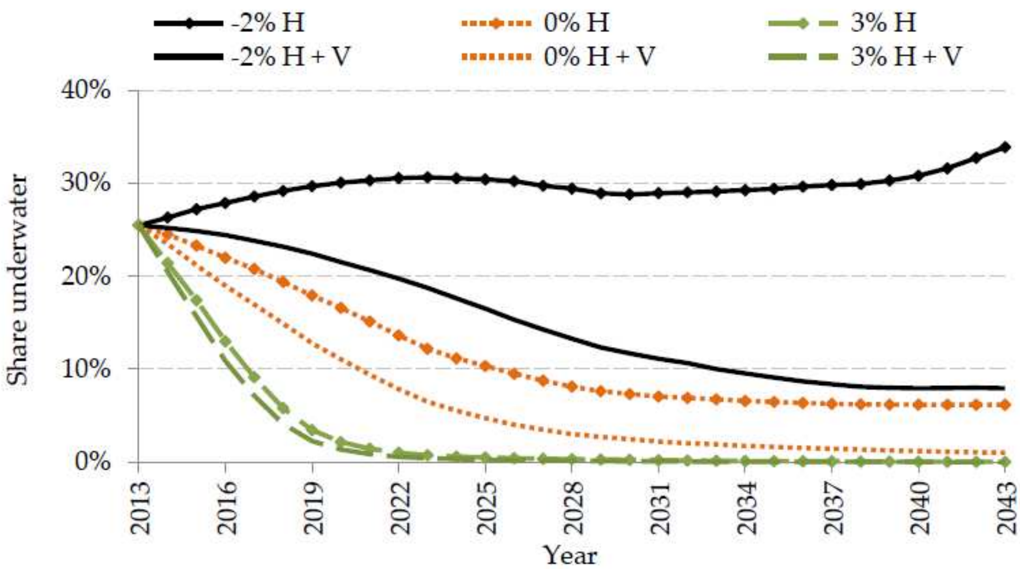



**Figure 9:** Simulation of the share of underwater mortgages among the mortgages currently existing in the

Netherlands, again considering different scenarios (H = house price change; V = voluntary repayments).

We will now concentrate on the mortgage characteristics at maturity, as most borrowers become no longer eligible to the tax-deductibility thereafter. To say something about the associated risk in terms of LGD we present the median home equity of all mortgages with the same maturity year in Figure 10. Moreover, from Figure 11 we observe that most mortgages mature around 2037, just like we expected from Figure 7. We find that the median home equity at maturity is positive for all years and in all scenarios, where only in the most pessimistic scenario the median home equity is close to zero in 2037. When house prices remain constant and voluntary repayments are allowed (which we consider to be the most realistic assumption), we find that only 3% of the mortgages that mature in 2037 are underwater.

It might be interesting to only focus on the mortgages that are currently underwater, as presented in Figure 12. From Figure 13 we observe that almost all of these mortgages are originated around 2008, as a result of the bursting of the housing bubble. For the mortgages that mature around this period, we observe a mean home equity that is again positive in almost all scenarios. Only in the most pessimistic scenario the median home equity is negative but close to zero. This is again explained by the fact that mortgages that are currently underwater typically contractually amortize at least to some extent, such that most of them are again above water at maturity.

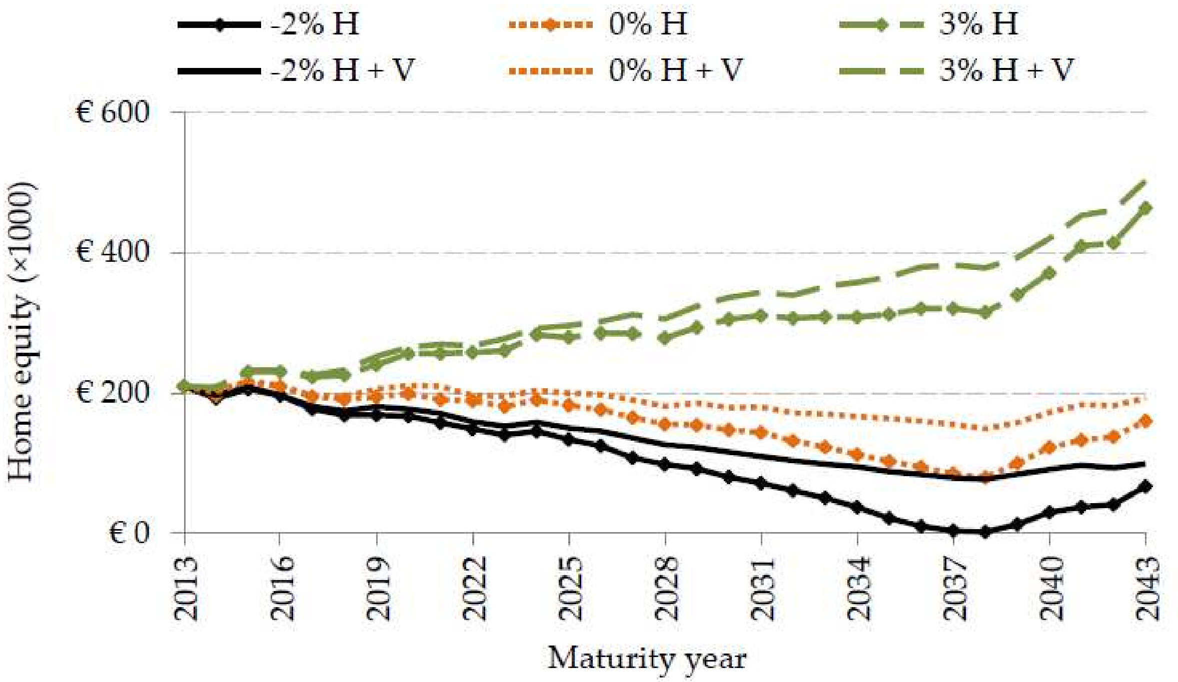



**Figure 10:** Median home equity per maturity year of the mortgages currently existing in the Netherlands.

Different scenarios are considered in the simulation (H = house price change; V = voluntary repayments).

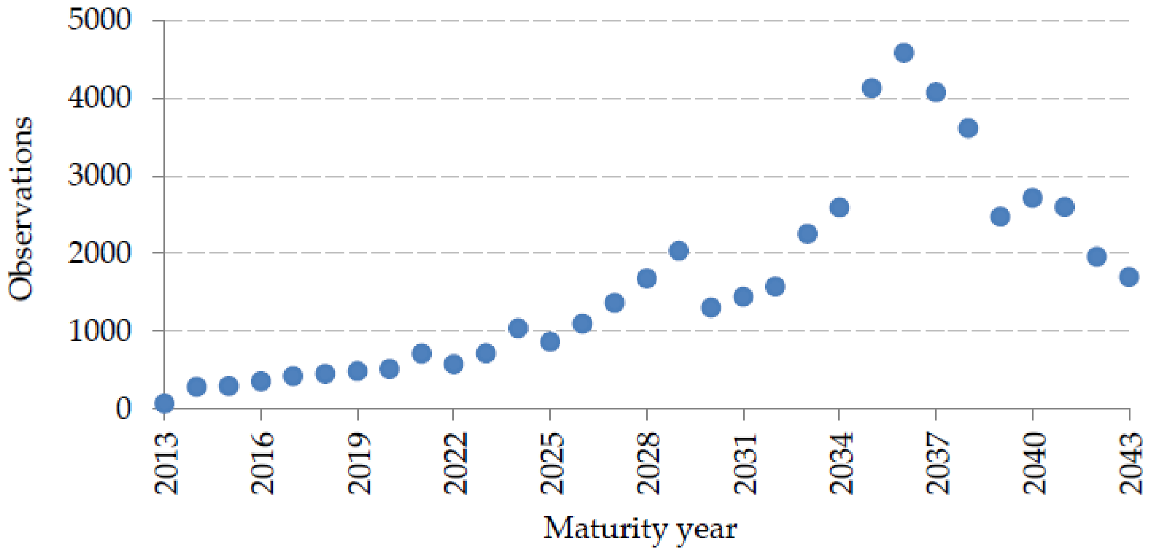



**Figure 11:** Number of borrowers in the simulation per maturity year of the corresponding mortgage.

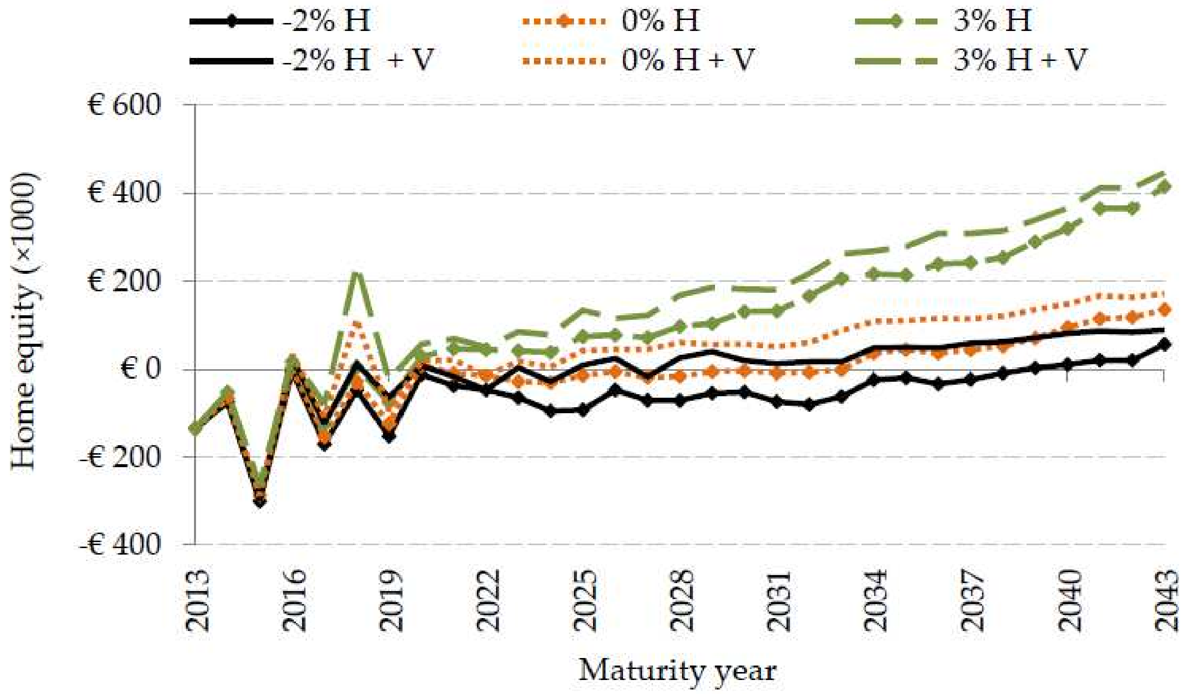



**Figure 12:** Median home equity per maturity year of the mortgages that are underwater in 2013. Different scenarios are considered (H = house price change; V = voluntary repayments).

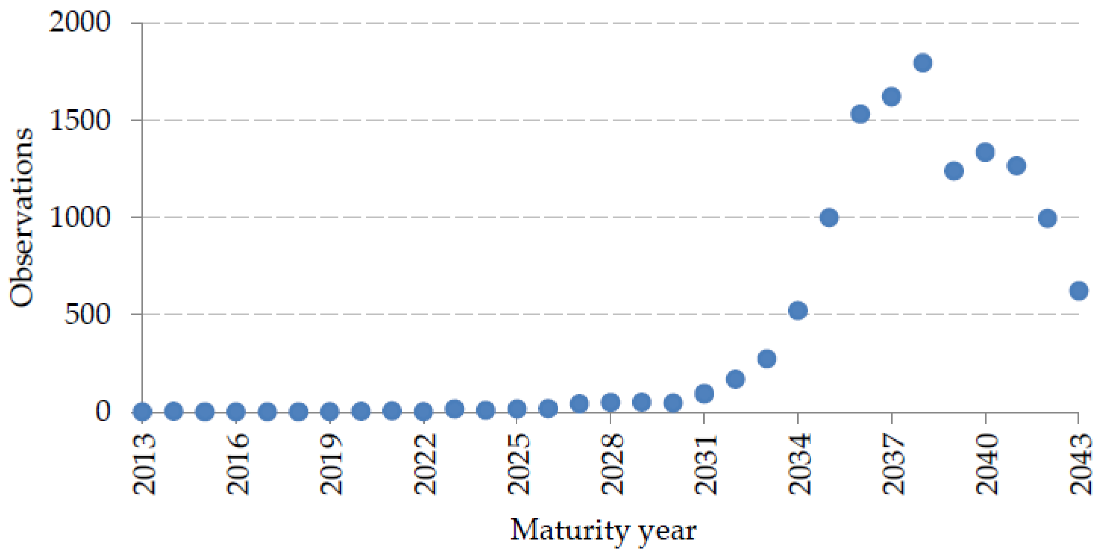
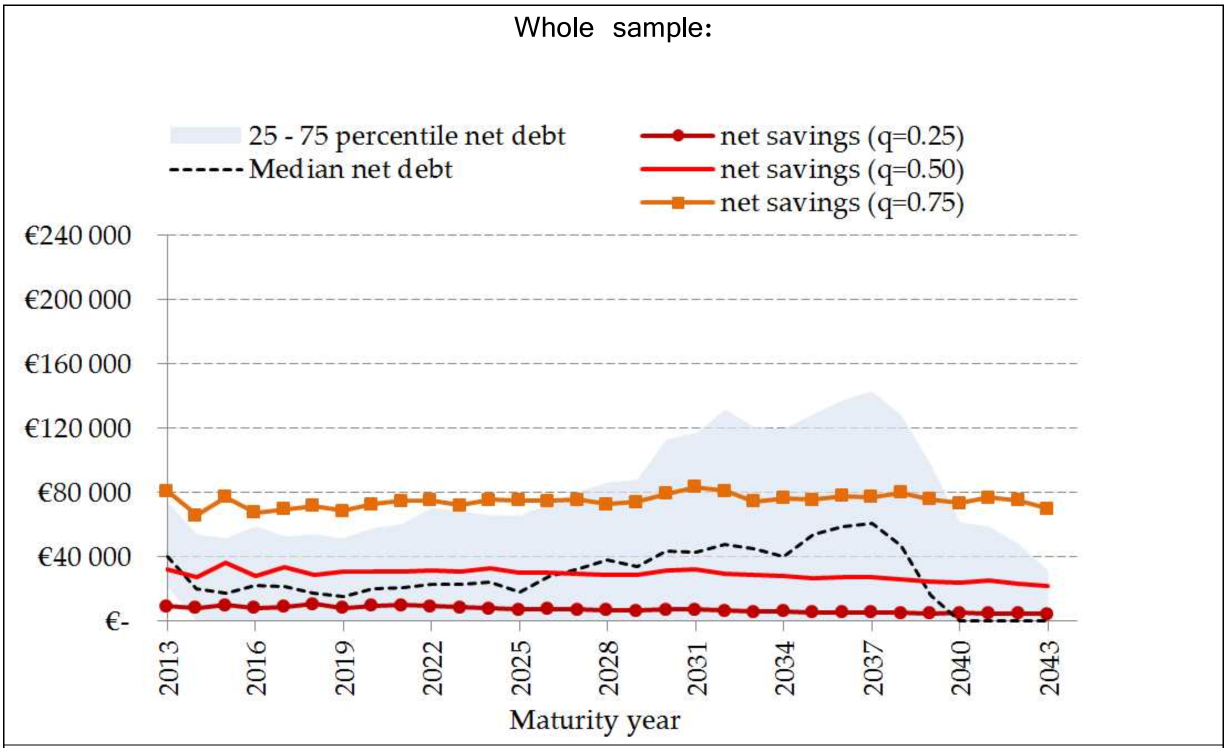



**Figure 13:** Number of borrowers in the simulation that are underwater in 2013 per maturity year of the corresponding mortgage.



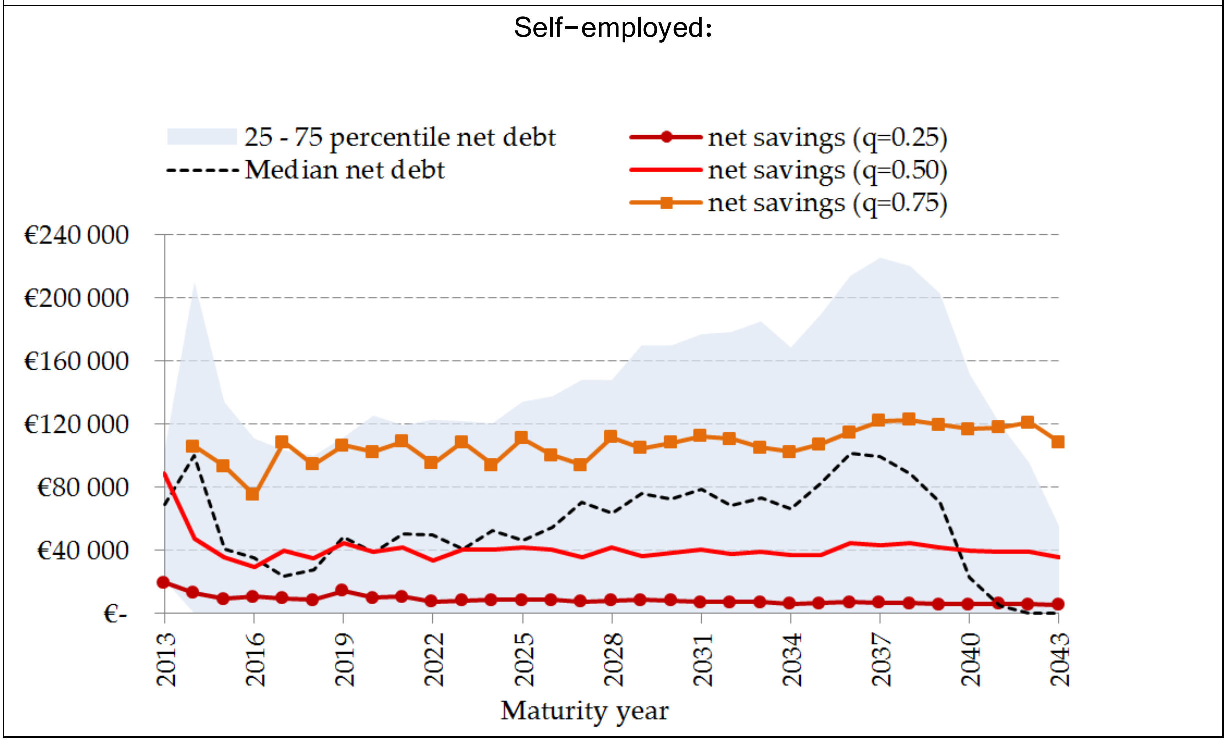




**Figure 14:** Distribution of net mortgage debt and net savings per maturity year. The quantiles of net savings represent the average of all estimated conditional quantiles of the borrowers with corresponding maturity year. Voluntary repayments are considered, house prices are assumed to remain constant and both GDP and CPI increase with 2% annually.

The final step in our analysis of the simulated data incorporates the distribution of net savings. Here we are interested in whether households have saved enough at maturity to fully repay their mortgage. Figure 14 presents the distribution of both net mortgage debt and net savings per maturity year of the whole sample and of self-employed. We only consider the scenario where house prices remain constant, as both net savings and net mortgage debt appear to be not very sensitive to house price changes in our model results. Interpreting the figure is rather difficult. We do not directly observe net savings of each borrower but only have estimations of the conditional expectation and quantiles. The quantiles of net savings presented in the figure represent the average of all conditional quantiles of the borrowers corresponding to a specific maturity year, which is not necessarily the same as the quantile of the distribution. Furthermore, the difference between median net savings and median net debt is not necessarily equal to the median of this difference. The figure does, however, provide the general impression that most borrowers will not have saved enough in order to repay debt at maturity, especially for those mortgages that get to mature in the period between 2030 and 2038. The figure also suggests that the heterogeneity in the debt distribution across self-employed is much larger than that of the whole sample.

Figure 15 presents the sensitivity of the distribution of net savings to different assumptions on the annual change in CPI and GDP. Especially the right tail of the distribution of net savings appears rather sensitive to different assumptions about CPI and GDP, whereas other parts of the distribution are not.

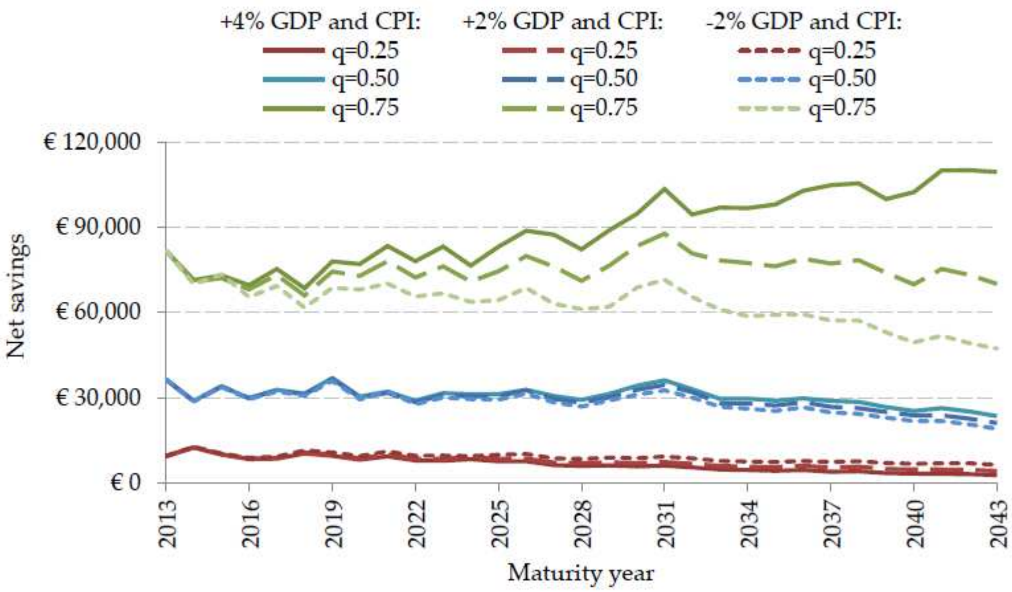



**Figure 15:** Sensitivity analysis of the distribution of net savings, where different CPI and GDP scenarios

are considered in the simulation. The quantiles represent the average of all estimated conditional quantiles of the borrowers with corresponding maturity year.

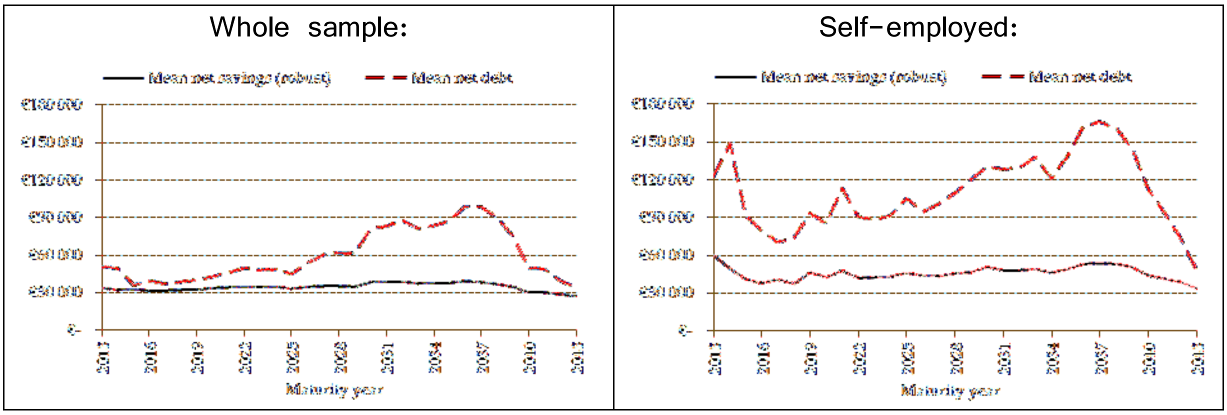



**Figure 16:** Average net debt and average net savings per maturity year. Here, average net savings are calculated by taking the average of the conditional expectations of all borrowers having a specific maturity year. House prices are considered to remain constant and both GDP and CPI increase with 2% annually.

Figure 16 presents the average mortgage debt and the average conditional expectation of net savings per maturity year. From this figure we indeed observe that on average, households will not have saved enough to repay the mortgage at maturity in all years. Mortgages that mature in the period between 2030 and 2038 will fall short of roughly 60 000 euro on average (100 000 for self-employed). They will have then saved about 30 000 euro (60 000 for the self-employed). This figure shows that self-employed are potentially more exposed to housing market risk, relative to the whole population.

This is even more evident in Figure 17, where we look at outstanding debt at the age when the mortgage interest deduction will expire (here computed as 30 years after origination). Only about 25% of the sample will be younger than 65 at maturity. In the figure, we also plot two lines representing the cumulative distribution of the share of the population whose mortgage matures by that age. Outstanding debt of self-employed in the ages between 65 and 70 (very likely future retirement ages), is higher for those who were self-employed at origination, while their savings are not (not shown in this figure). The figure shows that self-employed's outstanding debt is about twice as large, averaging sometimes at about 150 000 – 200 000 euro.

A back of the envelope computation suggests that, with an interest rate of 3%, the mortgage costs of those prolonging their loans after maturity would amount to monthly payments between 375-500 euro, if the self-employed could buy or keep an IO perpetuity (so he/she would continue to pay the interest rate only till death). However, not all IO loans have been sold as perpetuities. Given the old age of most respondents around maturity, it is plausible that banks would offer, after maturity, an annuity to be redeemed shortly, in say 10 years, for the residual part of debt. If this happens, the self-employed monthly costs would increase to about 1500-2000 euro.[10] For the non-self-employed, with an average outstanding debt of about 75 000 euro, the majority of the sample here, the monthly costs of a perpetuity or of a 10 years annuity vary between 190-740 euro a month.

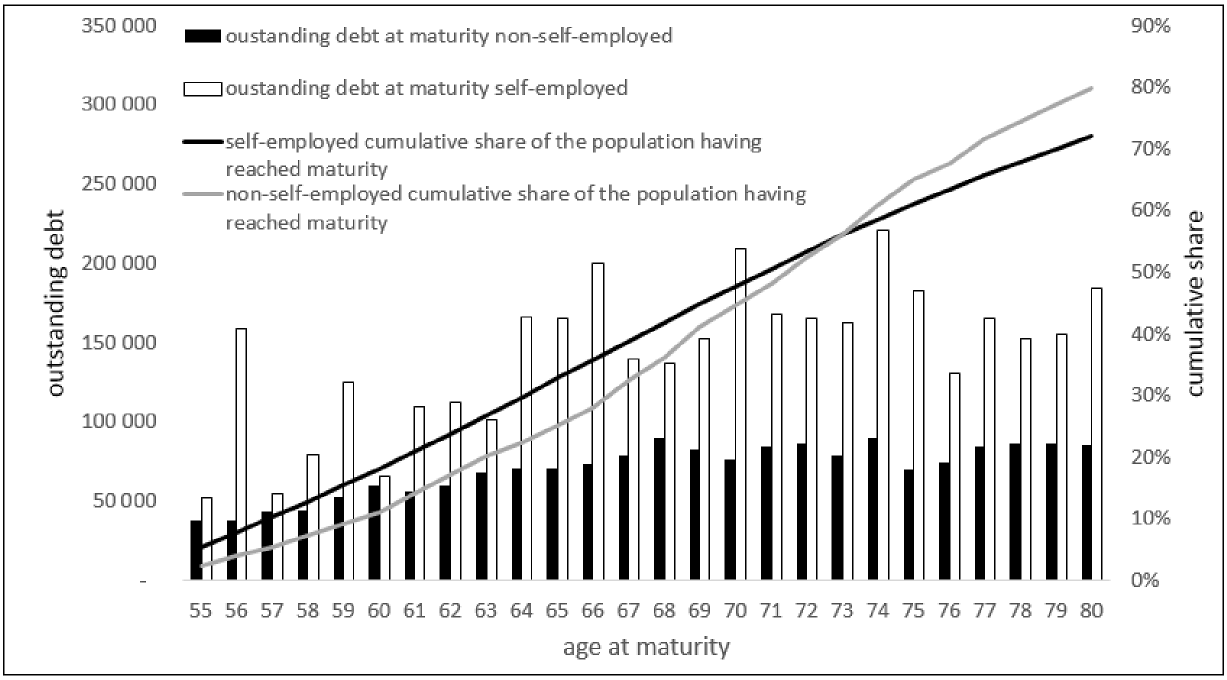



**Figure 17:** Outstanding mortgage debt at maturity by age of the mortgagor.

These computations show that the future financial burden of retirees with outstanding IO debt, will vary depending on their characteristics (self-employed being more indebted), institutions (mortgage interest deduction expiring), but also depending on the behavior of banks. Should banks offer IO perpetuities again, then the monthly costs of all households will be easily covered by the current social security benefit. Should this not be the case, say that banks were to offer a 10 years annuity mortgage, then the financial burden might become difficult to bare for the household. In the worst case depicted here, a self-employed household with a 200 000 euro outstanding debt shifting to a IO loan before maturity, to a 10-year annuity after, would experience an increase in monthly costs from 500 to 2000 euro, with a 3% interest rate.

[10] Assuming that private savings are not used to repay debt and assuming that the mortgage interest deduction is no longer available

## 5.3 Investment loans

In the computations above, we have not discussed the evolution of investment loans. It is fiscally unattractive and costly to redeem these products before maturity, so we have assumed that no voluntary repayments are observed on these products. Also, we expect that the pledged savings (BEW) in these loans are low. This expectation is due to the fact that the premiums paid in the saving/investment fund are low as they have been established assuming a high expected return (typically 8%). This mean expected return of 8% has not been achieved in the last decade, and even if it was, the costs and insurance premiums applied to these products might have prevented the pledged savings to repay debt at maturity. This implies that unrealistically high returns may be needed to achieve complete amortization of these loans. In addition, investment loans represent only about 5-7% of outstanding debt, so also at macro level the BEW's represent a relatively small cumulated capital.

[table 8 here]

In the data, we further observe that investment loans are in many cases only a part of the household mortgage. However, this does not mean that the residual part will provide any amortization. Table 8 shows that investment loans are very often combined with interest-only loans. For instance, 90% of those with an investment component of 60% of total mortgage have a residual share of their mortgage (40%) in interest-only loans and no other amortization. This means that limited amortization is present in the mortgage of those who own an investment loan.

# 6 Summary & conclusions

In this study, we show what part of the current mortgage debt has already been repaid in the Netherlands and what part is likely to be repaid before maturity, even if this debt is partly in interest-only mortgages or investment loans. Using a novel dataset, containing rich information on individual loan characteristics, we are able to shed light on the accumulated assets pledged to the mortgage and on voluntary repayments, two variables that are not observed in official statistics. We find that 58% of the current net mortgage debt comes from interest-only loans, but that these are often combined with amortizing loans. This is the case also for different subgroups, e.g. both for the wage-employed and self-employed. Borrowers having a full interest-only mortgage are typically older borrowers having substantial home equity, such that the risks regarding these mortgages are limited. Starters almost never have a full interest-only mortgage. We build a microsimulation model that simulates the mortgage debt thirty years in the future. In spite of the large share of interest-only loans, we find that about 1/3 of the current mortgage debt will be redeemed due to voluntary repayments. Many more interest-only loans will also be redeemed if we take mortality into account, as most interest-only-loans are with older borrowers, however mortality will reduce current debt only further than 30 years in the future.

In this study, we relate mortgage debt to non-housing wealth, and show that most households with a residual debt will not save enough to fully repay the mortgage at maturity. Especially, mortgages originated around the bursting of the housing bubble will have a substantial remaining debt (approximately 60 000 euro on average, but about 100 000 euro for the self-employed), that is not fully compensated by household financial wealth (30 000 euro on average and 60 000 for the self-employed). When we look at these figures by age, we show that households who are about to retire could be confronted with an increase in monthly costs depending on whether the bank will offer again an IO perpetuity or will demand a quicker repayment, as interest payments after mortgage maturity will no longer be tax-deductible. Specific groups, such as the self-employed and the owners of investment loans could then be confronted with larger financial problems. For instance, the mean self-employed with no financial wealth, shifting from a fully IO mortgage before maturity to a 10-year annuity after, will face an increase in monthly costs that is as high as the current social security benefit. Nevertheless, almost all borrowers will have a positive home equity at maturity, such that the risks associated to the banking sector will be limited.

## References


Almeida, H., M. Campello, and C. Liu (2006). The Financial Accelerator: Evidence from International Housing Markets. Review of Finance 10(3), 321–352.

Bernanke, B., M. Gertler, and S. Gilchrist (1996). The Financial Accelerator and the Flight to Quality. Review of Economics and Statistics 48(1), 1–15.

Bostic, R., S. Gabriel, and G. Painter (2009). Housing Wealth, Financial Wealth, and Consumption: New Evidence From Micro Data. Regional Science and Urban Economics 39(1), 79–89.

Bovenberg, Lans and Bas Jacobs (2008). Human capital and optimal positive taxation of capital income, Netspar DP 12/2008-056

Cameron, A. C. and P.K. Trivedi (2005). Microeconometrics: Methods and Applications. New York, NY: Cambridge University Press.

Campbell, J. Y. and J. F. Cocco (2007). How Do House Prices Affect Consumption? Evidence From Micro Data. Journal of Monetary Economics 54(3), 591–621.

Carson, R. T. and Y. Sun (2007). The Tobit Model with a Non-Zero Threshold. The Econometrics Journal 10(3), 488–502.

CBS (2014a). Documentatierapport Inkomenspanel Onderzoek (IPO) 2011. Voorburg: Centraal Bureau voor de Statistiek.

CBS (2014b). Documentatierapport Selectie Inkomenspanelonderzoek JJJJ Uit Integraal Vermogenbestand JJJJ+1 (Selectie IPO JJJJ Integraal-Vermogensbestand JJJJ+1). Voorburg: Centraal Bureau voor de Statistiek.

Cleveland, W. S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of the American Statistical Association 74(368), 829–836.

CPB (2014). CPB Financial Stability Report 2014. CPB Communication 2014, CPB Netherlands Bureau for Economic Policy Analysis.

Cragg, J. G. (1971). Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods. Econometrica 39(5), 829–844.

Dillingh, Rik, Henriette Prast, Mariacristina Rossi and Cesira Urzì Brancati (2015), The psychology and economics of reverse mortgage attitudes. Evidence from the Netherlands, Netspar Design Paper 38

DNB (2014). Overview of Financial Stability - Spring 2014. Amsterdam: Dutch Central Bank.

D'Orazio, M., M. Di Zio, and M. Scanu (2006). Statistical Matching: Theory and Practice. Chichester: Wiley.

Heckman, J. (1976). The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models. Annals of Economic and Social Measurement 5(4), 475–492.

Jansen, J., M. Bijlsma, M. Kruidhof, and C. Pattipeilohy (2013). Funding problems in the mortgage market. DNB Occasional Studies 11(1).

Kiyotaki, N. and J. Morre (1997). Credit Cycles. Journal of Political Economy 105(2), 211–248.

Koenker, R. (2005). Quantile Regression. New York, NY: Cambridge University Press.

Lorenzoni, G. (2008). Inefficient Credit Booms. Review of Economic Studies 75(3), 809–833.

Mastrogiacomo, M. and R. van der Molen (2015). Dutch mortgages in the DNB loan level data. DNB Occasional Studies, 4.

Mian, A and A Sufi (2015), House of debt: How they (and you) caused the Great Recession, and how we can prevent it from happening again, University of Chicago Press

Rässler, S. (2002). Statistical Matching: A Frequentist Theory, Practical Applications and Alternative Bayesian Approaches. New York, NY: Springer.

Sun, Wei and Triest, Robert K. and Webb, Anthony (2007). Optimal Retirement Asset Decumulation Strategies: The Impact of Housing Wealth (January 20, 2007). FRB of Boston Public Policy Discussion Paper No. 07-2.

Tobin, J. (1958). Estimation of Relationships For Limited Dependent Variables. Econometrica 26(1), 24 – 36.

Verardi, V. and C. Croux (2009). Robust Regression in Stata. The Stata Journal 9(3), 439–453.

Verbruggen J., R. van der Molen, S. Jonk, J. Kakes, W. Heeringa, (2015): Effecten van een verdere verlaging van de LTV-limiet, DNB Occasional Study (2)

Wooldridge, J. M. (2010). Econometric Analysis of Cross Section and Panel Data. Cambridge, MA: MIT Press.

# Tables

**Table 1:** Percentage of borrowers having a specific mortgage composition as reported in three waves of the LLD. Also, the share of each loan type at loan-level is presented, together with the total number of observations on both borrower- and loan-level.

| | 2012 Q4 | | 2013 Q3 | | 2013 Q4 | |
|---|---|---|---|---|---|---|
| Mortgage composition | borrowers | loans | borrowers | loans | borrowers | loans |
| *One loan type only* | | | | | | |
| Annuity | 1.35% | 3.55% | 1.98% | 4.58% | 2.36% | 5.12% |
| Linear | 0.61% | 0.98% | 0.70% | 1.09% | 0.72% | 1.13% |
| I-O | 35.90% | 60.99% | 35.46% | 59.59% | 37.06% | 60.34% |
| Savings | 6.90% | 15.52% | 7.32% | 16.45% | 6.79% | 15.59% |
| Life insurance | 4.63% | 11.15% | 4.53% | 10.22% | 4.19% | 9.59% |
| Investment | 3.66% | 5.52% | 2.97% | 4.84% | 2.31% | 4.49% |
| Other | 0.18% | 2.01% | 1.01% | 1.96% | 1.15% | 2.32% |
| Unknown | 0.71% | 0.28% | 0.69% | 1.28% | 0.78% | 1.42% |
| *Combination of loans* | | | | | | |
| Including IO | 44.98% | - | 44.03% | - | 43.08 | - |
| Excluding IO | 1.08% | - | 1.31% | - | 1.32 | - |
| Total observations | 3 040 976 | 5 828 982 | 2 928 214 | 5 641 773 | 2 915 542 | 5 611 558 |
| Total population (CBS) | 3 567 000 | | 3 562 500 | | 3 561 000 | |
| Coverage | 85.25% | | 82.20% | | 81.87% | |
| Reporting institutions | 7 | | 11 | | 9 (preliminary) | |

**Table 2:** Different property valuation methods used in the LLD (2013Q4)

| *Valuation method* | *Share* | *Property value* *Median* | *Mean* | *Std. Dev* |
|---|---|---|---|---|
| Internal and external expert inspection | 46.63% | 203 168 | 249 290 | 168 512 |
| External expert inspection only | 5.40% | 198 592 | 225 027 | 110 522 |
| Drive-by/desktop | 0.01% | 391 815 | 541 098 | 523 400 |
| Estate agent | 14.44% | 209 785 | 261 502 | 202 092 |
| WOZ-value | 17.52% | 225 979 | 257 842 | 151 608 |
| Other/unknown | 16.00% | 242 084 | 318 079 | 232 931 |

**Table 3:** Descriptives LLD 2013 Q4 on borrower-level per I-O category

| | 0%IO | | | 20%IO | | | 40%IO | | |
|---|---|---|---|---|---|---|---|---|---|
| *Variable* | *Mean* | *Std. dev.* | *Median* | *Mean* | *Std. dev.* | *Median* | *Mean* | *Std. dev.* | *Median* |
| Age | 45.4 | 12.5 | 45.0 | 44.4 | 9.6 | 44.0 | 41.5 | 10.0 | 41.0 |
| House value ( ) | 235 537 | 159 719 | 200 994 | 249 325 | 140 356 | 215 214 | 226 241 | 131 914 | 196 399 |
| Net debt ( ) | 146 662 | 125 367 | 133 251 | 192 079 | 172 190 | 119 908 | 198 454 | 111 618 | 180 332 |
| LTV (%) | 68 | 42 | 75 | 81 | 33 | 85 | 93 | 32 | 103 |
| Interest rate (%) | 4.6 | 1.1 | 4.7 | 4.7 | 0.8 | 4.7 | 4.6 | 0.7 | 4.7 |
| NHG (%) | 38 | | | 35 | | | 54 | | |
| Underwater (%) | 30 | | | 33 | | | 54 | | |
| Observations | 535 830 | | | 104 323 | | | 314 786 | | |
| | 60% I-O | | | 80% I-O | | | 100% I-O | | |
| *Variable* | *Mean* | *Std. dev.* | *Median* | *Mean* | *Std. dev.* | *Median* | *Mean* | *Std. dev.* | *Median* |
| Age | 46.3 | 10.0 | 46.0 | 51.3 | 10.9 | 51.0 | 60.4 | 12.3 | 61.0 |
| House value ( ) | 261 968 | 159 964 | 221 530 | 292 854 | 195 023 | 242 008 | 300 081 | 212 823 | 247 548 |
| Net debt ( ) | 215 640 | 140 022 | 188 288 | 227 015 | 170 039 | 189 750 | 142 995 | 145 629 | 106 000 |
| LTV (%) | 86 | 32 | 92 | 81 | 34 | 83 | 48 | 30 | 44 |
| Interest rate (%) | 4.6 | 0.8 | 4.7 | 4.5 | 0.9 | 4.6 | 4.4 | 1.0 | 4.5 |
| NHG (%) | 26 | | | 10 | | | 4 | | |
| Underwater (%) | 38 | | | 34 | | | 5 | | |
| Observations | 323 206 | | | 204 976 | | | 892 425 | | |

**Table 4:** Descriptives IPO 2005, 2008 and 2011

| | 2005 | | | 2008 | | | 2011 | | |
|---|---|---|---|---|---|---|---|---|---|
| *Variable* | | | | | | | | | |
| | *Mean* | *Std. dev* | *Median* | *Mean* | *Std. dev* | *Median* | *Mean* | *Std. dev* | *Median* |
| Age | 45.3 | 11.7 | 43 | 46.2 | 11.9 | 44 | 47.8 | 11.9 | 46 |
| House value ( ) | 280 025 | 258 908 | 237 412 | 308 320 | 179 159 | 261 016 | 280 622 | 160 106 | 238 225 |
| Gross debt ( ) | 163 032 | 169 507 | 135 500 | 194 445 | 174 385 | 163 600 | 206 674 | 175 184 | 176 000 |
| Net savings ( ) | 44 292 | 227 898 | 18 808 | 39 755 | 285 174 | 18 642 | 38 750 | 270 909 | 18 171 |
| Interest rate (%) | 5.2 | 1.4 | 5.1 | 4.9 | 1.0 | 4.8 | 4.8 | 1.0 | 4.8 |
| Observations | 42 998 | | | 50 171 | | | 49 562 | | |

**Table 5:** Estimation results for different models for voluntary repayments (Tobit in levels, Tobit in logs and the Cragg log-normal hurdle)

| | Probit (Part I) | | Tobit (in levels) | | Tobit (in logs) | | Two-Part (Part II) |
|---|---|---|---|---|---|---|---|
| | Coef | ME | Coef | ME ($\Pr(y_i > 0 \mid \mathbf{x}_i)$) | Coef | ME ($\Pr(y_i > 0 \mid \mathbf{x}_i)$) | Coef |
| age/10 | 0.261*** | -0.00968*** | 21 170*** | -0.00357*** | 1.000*** | -0.00801*** | 0.447*** |
| | (0.00687) | (0.000241) | (448.6) | (0.000223) | (0.0233) | (0.000235) | (0.0127) |
| $(\mathrm{age}/10)^2$ | -0.0317*** | | -2354*** | | -0.118*** | | -0.0423*** |
| | 0.000619) | | (4.06) | | (0.00210) | | (0.00116) |
| share I-O | 0.209*** | 0.0452*** | 14,061*** | 0.0442*** | 0.741*** | 0.0468*** | 0.208*** |
| | (0.00337) | (0.000730) | (219.8) | (0.000688) | (0.0114) | (0.000718) | (0.00607) |
| interest rate | 1.125*** | 0.244*** | -44,469*** | -0.140*** | 1.225*** | 0.0774*** | -8.236*** |
| | (0.135) | (0.0292) | (8 777) | (0.0276) | (0.456) | (0.0288) | (0.260) |
| underwater | -0.673*** | -0.00660*** | -53,579*** | -0.00201* | -2.550*** | -0.00992*** | -0.912*** |
| | (0.0142) | (0.00116) | (912.2) | (0.00105) | (0.0480) | (0.00112) | (0.0263) |
| age × underwater | 0.0126*** | | 1,038*** | | 0.0469*** | | 0.0146*** |
| | (0.000332) | | (21.06) | | (0.00111) | | (0.000603) |
| NHG | -0.0985*** | -0.0214*** | -4,677*** | -0.0147*** | -0.354*** | -0.0224*** | -0.0844*** |
| | (0.00327) | (0.000709) | (211.2) | (0.000663) | (0.0110) | (0.000696) | (0.00593) |
| $\mathrm{currentLTV}/10^2$ | -0.123*** | -0.0267*** | 4709*** | 0.0148*** | -0.102*** | -0.00642*** | 0.914*** |
| | (0.00495) | (0.00107) | (320.0) | (0.00101) | (0.0167) | (0.00105) | (0.00877) |
| Constant | -1.597*** | | -122 037*** | | 1.732*** | | 8.204*** |
| | (0.0205) | | (1 343) | | (0.0698) | | (0.0375) |
| $N$ | 1 901 566 | | 1 901 566 | | 1 901 566 | | 1 901 566 |
| pseudo $R^2$ | 0.010 | | 0.006 | | 0.010 | | |
| Log-likelihood | -750 856 | | -3 712 000 | | -3667 000 | | -2881 000 |
| $\hat{\sigma}^2$ | | | 65 986 | | | | 1.048 |
| *Two-Part model:* | | | | | | | |
| pseudo $R^2$ | | | | | | | 0.010 |
| Log-likelihood | | | | | | | -3632000 |
| *p-value LM tests:* | | | | | | | |
| heteroskedasticity | 0.000 | | | | | | |
| normality | 0.000 | | | | | | |

Standard errors below coefficients *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

**Table 6:** Three probability models (linear, logit and probit) for the participation decision to voluntarily repay (1 = voluntary repayment, 0 = no voluntary repayment).

| | Linear Probability | | Logit | | Probit | |
|---|---|---|---|---|---|---|
| | Coef | ME | Coef | ME | Coef | ME |
| Age/10 | 0.0480*** | -0.00964*** | 0.494*** | -0.00994*** | 0.261*** | -0.00968*** |
| | -0.00143 | -0.000257 | -0.0131 | -0.000241 | -0.00687 | -0.000241 |
| $(Age/10)^2$ | -0.00610*** | | -0.0600*** | | -0.0317*** | |
| | -0.000127 | | -0.00118 | | -0.000619 | |
| Share I-O | 0.0462*** | 0.0462*** | 0.387*** | 0.0454*** | 0.209*** | 0.0452*** |
| | -0.00073 | -0.00073 | -0.00626 | -0.000734 | -0.00337 | -0.00073 |
| Interest rate | 0.234*** | 0.234*** | 2.143*** | 0.251*** | 1.125*** | 0.244*** |
| | -0.029 | -0.029 | -0.246 | -0.0289 | -0.135 | -0.0292 |
| Underwater | -0.119*** | -0.00654*** | -1.309*** | -0.00603*** | -0.673*** | -0.00660*** |
| | -0.00292 | -0.00114 | -0.0275 | -0.00116 | -0.0142 | -0.00116 |
| Age * underwater | 0.00222*** | | 0.0246*** | | 0.0126*** | |
| | -0.0000694 | | -0.000632 | | -0.000332 | |
| NHG | -0.0196*** | -0.0196*** | -0.182*** | -0.0214*** | -0.0985*** | -0.0214*** |
| | -0.000688 | -0.000688 | -0.00622 | -0.00073 | -0.00327 | -0.000709 |
| Current LTV / 100 | -0.0300*** | -0.0300*** | -0.237*** | -0.0278*** | -0.123*** | -0.0267*** |
| | -0.00109 | -0.00109 | -0.00913 | -0.00107 | -0.00495 | -0.00107 |
| Constant | 0.0534*** | | -2.796*** | | -1.597*** | |
| | -0.00433 | | -0.0385 | | -0.0205 | |
| N | 1901566 | | 1901566 | | 1901566 | |
| Pseudo R2 | 0.01 | | 0.01 | | 0.01 | |
| Log-likelihood | -760934 | | -750842 | | -750856 | |

Standard errors below coefficients *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

**Table 7:** Estimation results of a robust regression on net savings and three quantile regressions on the inverse hyperbolic sine transformation of net savings.

| | Robust regression (in levels) | Quantile regression (IHS transformed) q=0.25 | q=0.5 | q=0.75 |
|---|---|---|---|---|
| age/10 | 526.08** | 0.0947*** | 0.0750*** | 0.0802*** |
| | -265.13 | -0.00979 | -0.00584 | -0.00525 |
| (age/10)^2 | -324.6 | -0.0765*** | -0.0571*** | -0.0617*** |
| | -274.65 | -0.00983 | -0.00573 | -0.00566 |
| gross mortgage debt/10^3 | 13.979** | 0.000562** | 0.000537*** | 0.000328*** |
| | -5.924 | -0.00025 | -9.3E-05 | -9.7E-05 |
| property value/10^3 | -2.482 | -5.9E-05 | -0.00016 | -0.00016 |
| | -5.806 | -0.00032 | -0.00016 | -0.00011 |
| interest rate | 2698.6 | 0.526 | -0.396 | -1.083** |
| | -10458 | -0.616 | -0.285 | -0.431 |
| CPI | -4.262 | -0.00576 | 0.00157 | 0.00522*** |
| | -39.53 | -0.00377 | -0.00242 | -0.00194 |
| GDP | 25.181 | 0.00337 | -0.00102 | -0.00265 |
| | -34.79 | -0.00385 | -0.00253 | -0.00236 |
| # transactions per postcode/10^2 | -12.705** | 0.00348*** | 0.00202*** | 0.00107*** |
| | -26.85 | -0.00079 | -0.00049 | -0.00041 |
| I-O share per postcode | -6310.179*** | 0.846*** | -0.203*** | -0.500*** |
| | -4410 | -0.107 | -0.0685 | -0.0515 |
| mean house price per postcode/10^3 | -0.835* | 0.0000109** | 4.65E-06 | -9.2E-07 |
| | -2.49 | -4.7E-06 | -3.8E-06 | -2.9E-06 |
| Average gross mortgage debt/10^2 | -3.447*** | -0.000334*** | -0.000165*** | -0.0000919*** |
| | -0.76 | -2.3E-05 | -1E-05 | -1.1E-05 |
| Average property value/10^2 | 7.275*** | 0.000351*** | 0.000413*** | 0.000444*** |
| | -0.77 | -3.4E-05 | -1.7E-05 | -1.2E-05 |
| Average interest rate | 107430*** | 11.40*** | 9.564*** | 9.054*** |
| | -28208 | -1.029 | -0.431 | -0.718 |
| Birth cohorts | Yes | Yes | Yes | Yes |
| Constant | -24043*** | 5.518*** | 7.507*** | 8.546*** |
| | -74470 | -0.308 | -0.288 | -0.156 |
| N | 341118 | 341118 | 341118 | 341118 |
| R2 | | 0.0137 | 0.0343 | 0.0578 |
| p-value Wald test for joint significance: | | | | |
| Birth cohorts | 0.927 | 0 | 0 | 0 |
| Postcode variables | 0.481 | 0 | 0 | 0 |

Standard errors in below coefficients; panel-robust bootstrap standard errors are reported for the robust regression *** p < 0.01, ** p < 0.05, * p < 0.1

**Table 8:** Combination of investment loans with interest-only loans

| | 20% investment | 40% investment | 60% investment | 80% investment | 100% investment |
|---|---|---|---|---|---|
| no interest-only component | 3% | 4% | 7% | 11% | 100% |
| 20% interest-only | 3% | 4% | 3% | 89% | |
| 40% interest-only | 9% | 7% | 90% | | |
| 60% interest-only | 12% | 85% | | | |
| 80% interest-only | 73% | | | | |

**Explanatory note:** The diagonal cells indicate no amortization. The residual category is non-investment and non interest-only loan.

# Appendix

## Voluntary repayments

Let $y_i$ denote the voluntary repayments for borrower $i = 1,2,\dots,N$. We have that $y_i$ takes on the value zero with positive probability, but is a continuous random variable over strictly positive values. Variables with this specific characteristic are typically modeled using corner solution response models (see Wooldridge (2010) for an introduction to corner response models). We will compare a number of different model specifications, where comparison is based on, among others, the log-likelihood and pseudo $R^2$. We use the squared correlation between fitted values and actual observations as a measure for the pseudo $R^2$, as they are directly comparable across classes of models. First, we consider a standard Tobit model (Tobin, 1958):

$$y_i^* = \mathbf{x}_{i\prime}\boldsymbol{\beta} + \varepsilon_i, i = 1,2,\dots,N, \quad (2)$$

In our specification we use the following $k = 8$ explanatory variables: age, age squared, current LTV, debt-weighted share of interest-only loans, mortgage interest rate, a dummy indicating the borrower has NHG, a dummy indicating the mortgage is underwater and an interaction term between age and the underwater dummy. Now, instead of observing the latent variable $y_i^*$, we observe

$$y_i = \begin{cases} y_i^* & \text{if} y_i^* \geq L \\ 0 & \text{if} y_i^* < L, \end{cases} \tag{3}$$

where we argued to set $L = 2000$. Maximum likelihood estimation of the standard Tobit model with zero censoring point is explained in standard econometric textbooks (e.g. Cameron and Trivedi (2005)). However, here we are dealing with a non-zero threshold. We estimate $\boldsymbol{\beta}$ by running a standard Tobit on $y_i^\bullet = \max(0, y_i^* - L)$, which has zero censoring point, and then adjust the estimated intercept by $L$.

We furthermore define the participation equation

$$w_i = \begin{cases} 1 & \text{if} y_i > 0 \\ 0 & \text{if} y_i = 0, \end{cases} \tag{4}$$

such that the conditional probability of a voluntary repayment is given by

$$\begin{aligned} \Pr(w_i = 1|\mathbf{x}_i) &= \Pr(y_i^* \geq L|\mathbf{x}_i) \\ &= \Pr(\mathbf{x}_{i'}\boldsymbol{\beta} + \varepsilon_i \geq L) \\ &= \Pr\left(\frac{\varepsilon_i}{\sigma} \geq \frac{L - \mathbf{x}_{i'}\boldsymbol{\beta}}{\sigma}\right) \\ &= \Phi\left(\frac{\mathbf{x}_{i'}\boldsymbol{\beta} - L}{\sigma}\right), \end{aligned}$$

where the last step follows since the distribution of $\varepsilon_i$ is symmetric around zero. Hence, if (2) and (3) are true, $w_i$ follows a probit model. By running a probit model on $w_i$, we can test for heteroskedasticity and normality in the error term of the latent equation (2). The probit and Tobit should yield similar parameter estimates, as they are based on the same latent model. Notice, however, that $\sigma$ and $\boldsymbol{\beta}$ are not uniquely identified in a probit model (for identifiability, it is assumed that $\sigma = 1$). Instead, we get an estimate of the $(k+1) \times 1$ vector $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_{k+1})' = ((\beta_1 - L)/\sigma, \beta_2/\sigma, \beta_3/\sigma, \dots, \beta_{k+1}/\sigma)$. Some manipulations of the Tobit estimates are therefore necessary to make them comparable with the probit estimates. As $\sigma > 0$, we would at least expect that Tobit and probit estimates have the same sign. One could also compare the marginal effects (ME) of a change in regressor on $\Pr(y_i > 0|\mathbf{x}_i)$ with the ME from the probit model. Let $x_{ij}$ denote the $j$th component of $\mathbf{x}_i$. Now, the ME of change in regressor $x_{ij}$ on $\Pr(y_i > 0|\mathbf{x}_i)$ is given by

$$\frac{\partial \Pr(y_i>0|\mathbf{x}_i)}{\partial x_{ij}} = \frac{\beta_j}{\sigma}\phi\left(\frac{\mathbf{x}_i\prime\boldsymbol{\beta}-L}{\sigma}\right), \qquad (5)$$

for $j = 2, \dots, k+1$. Also, the ME for the probit model are given by

$$\frac{\partial \Pr(y_i>0|\mathbf{x}_i)}{\partial x_{ij}} = \gamma_j\phi(\mathbf{x}_i\prime\boldsymbol{\gamma}),$$

which is the same as (5) (notice that the ME for $j = 1$ is not considered, as $x_{i1}$ is a constant). Altogether, the estimated ME resulting from the Tobit estimates should be similar to the ME from the probit model if the Tobit model is correctly specified.

We observed that the distribution of the voluntary repayments was highly right-skewed with considerable non-normal kurtosis. It might work better to take the natural logarithm. Now, instead of (2) and (3) we introduce a log-normal variant of the standard Tobit model by specifying

$$y_i^* = \exp(\mathbf{x}_i\prime\boldsymbol{\beta} + \varepsilon_i), \varepsilon_i|\mathbf{x}_i \sim NID(0, \sigma^2)$$

where we should note that $\boldsymbol{\beta}$, $\varepsilon_i$ and $\sigma^2$ are redefined and not the same as in (2). Moreover, we observe

$$y_i = \begin{cases} y_i^* & \text{if } \ln(y_i^*) \geq \ln(L) \\ 0 & \text{if } \ln(y_i^*) < \ln(L). \end{cases}$$

Notice that $\ln(0)$ is not defined, such that all censored observations are lost when transforming to log-normal data. Among others, Carson and Sun (2007) show that consistent estimates are obtained by setting all censored observations to the minimum non-censored value of $\ln y_i$.[11]

The Tobit model has some restrictive implication, e.g. the ME of $x_{ij}$ on $\Pr(y_i > 0|\mathbf{x}_i)$ and $\mathrm{E}(y_i|\mathbf{x}_i, y_i > 0)$ always have the same sign. By relaxing these assumptions we might obtain a better fit. Thus we consider the Cragg log-normal hurdle (Cragg, 1971), or Two-Part model, which allows separate mechanisms to determine the participation decision ($w_i = 0$ or $w_i = 1$) and the amount decision (magnitude of $y_i$ when $y_i > 0$). Here we express $y_i$ as follows:

$$y_i = w_i \cdot y_i^* = I(\mathbf{x}_i\prime\boldsymbol{\lambda} + v_i > L)\exp(\mathbf{x}_i\prime\boldsymbol{\delta} + u_i), \qquad (6)$$

[11] Actually, when using a canned statistical package like STATA, we need to set the censored observations to an amount slightly smaller than the minimum non-censored value of $\ln y_i$ (i.e. $\ln(L) - 1.10^{-6}$). Otherwise, the minimum non-censored value will be treated as a censored value as well.

where $I(.)$ is the indicator function, $v_i|\mathbf{x}_i \sim NID(0,1)$ and $u_i|\mathbf{x}_i \sim NID(0,\sigma^2)$ and where we assume $v_i$ and $u_i$ are independent. As can be seen, the same regressors are used in both parts, as there are no obvious exclusion restrictions. Estimation is done in two parts. First, we run a probit regression on $w_i$ to estimate $\boldsymbol{\lambda}$ (Part I). Second, we estimate $\boldsymbol{\delta}$ and $\sigma^2$ by running an OLS regression on $\ln y_i$ using only the observations for which $y_i > 0$ (Part II).

The assumption that $v_i$ and $u_i$ are independent might be rather strong. The Heckman selection model (Heckman, 1976) relaxes this independence assumption. However, identification of such a model can be fragile without a valid exclusion restriction, i.e. a variable that affects the selection equation but not the main equation. It is hard to find such a variable in practice. Moreover, for practical reasons we also choose not to consider a Heckman model; a Cragg log-normal hurdle is much easier to implement in the simulation.

### Non-housing wealth

Let $y_{it}$ denote the net savings for borrower $i$ at time $t$. The distribution of net savings is highly right-skewed and can have both extreme positive and negative values. Using the natural logarithm to normalize the distribution of the data does not help, as log-transformations for non-positive observations are not defined. Keeping this in mind, let us consider the following panel model:

$$y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + c_i + u_{it}, i = 1, \ldots, N; t = 1, \ldots, T. \tag{7}$$

where $c_i$ is an unobserved individual effect, $u_{it}$ is an error term and $\mathbf{x}_{it}$ is a $(k+1)\times 1$ vector including $k$ regressors and a constant. Here, we assume the observations are independent across individuals, but not necessarily across time. Regarding the error term, we only make the assumption that $\mathrm{E}(u_{it}|\mathbf{x_{it}}, c_i) = 0$. Hence, for reasons discussed above, we do not make the usual assumptions that $u_{it}$ is i.i.d. and normally distributed. Moreover, we assume $\mathrm{E}(c_i|\mathbf{x}_i) = 0$, where $\mathbf{x}_i = (\mathbf{x}_{i1}', \ldots \mathbf{x}_{iT}')'$. If we make the fixed effect assumption instead, i.e. $\mathrm{E}(c_i|\mathbf{x}_i) \neq 0$, we cannot estimate $c_i$ for the individuals in the LLD (estimation of the individual-specific effect requires that net savings are observed in at least one time period for that specific individual). Instead, we try to imitate fixed effects by including a number of time-invariant regressors in $\mathbf{x}_{ij}$. In total, we use the following $k = 28$ regressors: age, age squared, gross mortgage debt, property value, mortgage interest rate, nominal consumer price index (CPI), nominal gross domestic product (GDP), three variables on postcode-level (number of real estate transactions, average debt-weighted share of interest-only mortgage and average property value), three time-invariant variables constructed by averaging time-varying variables over time (average gross mortgage debt, average property value and average interest rate) and fifteen cohort dummies.

Now, let $v_{it} = c_i + u_{it}$ such that (7) can be rewritten as $y_{it} = \mathbf{x}_{it}'\boldsymbol{\beta} + v_{it}$. The assumptions on $u_{it}$ and $c_i$ imply that $\mathrm{E}(v_{it}|\mathbf{x}_{ij}) = 0$, such that the conditional expectation of $y_{it}$ is given by $\mathrm{E}(y_{it}|\mathbf{x_{it}}) = \mathbf{x}_{it}'\boldsymbol{\beta}$. $\mathrm{E}(v_{it}|\mathbf{x}_{ij}) = 0$ is sufficient to prove that $\beta$ can be consistently estimated using Pooled OLS. However, simple OLS regression is highly sensitive to the presence of outliers in the data and might be inefficient under highly non-normal errors. To deal with this, several robust regressions have been proposed in the literature, yielding a more resistant estimate of $\boldsymbol{\beta}$. The general idea is that most influential observations in the simple OLS regression (associated with Cook's distances larger than one) are dropped, after which the remaining observations with large absolute residuals are down-weighted. The exact down-weighting procedure for the specific robust regression we use in this study is extensively described in Verardi and Croux (2009). Now, let the estimate of $\boldsymbol{\beta}$ resulting from the robust regression be denoted by $\widehat{\boldsymbol{\beta}}$. To obtain panel-robust standard errors we apply the bootstrap method. Specifically, $B = 50$ pseudo-samples of $N_b = 10000$ borrowers are constructed by drawing with replacement over $i$ and using all observed time periods for that borrower. For each pseudo-sample, we perform a robust regression of $y_{it}$ on $\mathbf{x}_{it}$, yielding $B$ estimates of $\boldsymbol{\beta}$ denoted by $\widehat{\boldsymbol{\beta}}_b, b = 1, \dots, B$. Now, let $\overline{\widehat{\boldsymbol{\beta}}} = \frac{1}{B}\sum_{b=1}^{B} \widehat{\boldsymbol{\beta}}_b$, such that the panel bootstrap estimate of the variance matrix of $\widehat{\boldsymbol{\beta}}$ is given by

$$\widehat{\mathbf{V}}_{boot}(\widehat{\boldsymbol{\beta}}) = \frac{1}{B-1}\sum_{b=1}^{B} (\widehat{\boldsymbol{\beta}}_b - \overline{\widehat{\boldsymbol{\beta}}})(\widehat{\boldsymbol{\beta}}_b - \overline{\widehat{\boldsymbol{\beta}}})'.$$

Next, quantile regression (QR) is used to provide a more complete picture of the conditional distribution of $y_{it}$. In contrast to OLS regression, QR is robust against outliers and is equivariant to monotone transformations. This last property is important here, as we need to transform the data in order to achieve convergence in the quantile regression. Specifically, we apply the inverse hyperbolic sine (IHS) transformation to $y_{it}$:

$$y_{it}^{\bullet} = \sinh^{-1}(y_{it}) = \ln\left(y_{it} + \sqrt{y_{it}^2 + 1}\right),$$

where the hyperbolic sine function is used to transform the data back:

$$y_{it} = \sinh(y_{it}^{\bullet}) = \frac{1}{2}\left(e^{y_{it}^{\bullet}} - e^{-y_{it}^{\bullet}}\right).$$

Now, let $q \in (0,1)$ and denote the $q$th conditional quantile of the distribution of $y_{it}^{\bullet}$ by $Q_q(y_{it}^{\bullet}|\mathbf{x}_{it})$, where we assume $Q_q(y_{it}^{\bullet}|\mathbf{x}_{it})$ is linear in $\mathbf{x}_{it}$, i.e. $Q_q(y_{it}^{\bullet}|\mathbf{x}_{it}) = \mathbf{x}_{it}'\boldsymbol{\beta}_q$. The subscript in $\boldsymbol{\beta}_q$ indicates that the parameters are different for different points in the conditional distribution. In particular, we estimate $\boldsymbol{\beta}_q$ for $q = 0.25, 0.50, 0.75$. Estimation of $\boldsymbol{\beta}_q$ is done by minimizing the following objective function:

$$Q_N(\boldsymbol{\beta}_q) = \sum_{i:y_{it}^{\bullet} \geq \mathbf{x}_{it}'\boldsymbol{\beta}_q}^{N} q|y_{it}^{\bullet} - \mathbf{x}_{it}'\boldsymbol{\beta}_q| + \sum_{i:y_{it}^{\bullet} < \mathbf{x}_{it}'\boldsymbol{\beta}_q}^{N} (1-q)|y_{it}^{\bullet} - \mathbf{x}_{it}'\boldsymbol{\beta}_q|.$$

This objective function is not differentiable, but fortunately linear programming methods can be used to solve the minimization problem (see Koenker (2005)). After obtaining an estimate for $Q_q(y_{it}^{\bullet}|\mathbf{x}_{it})$, we simply transform this estimate using the hyperbolic sine function to get an estimate for $Q_q(y_{it}|\mathbf{x}_{it})$. Again, the bootstrap method should be used to obtain panel-robust standard errors, which adds considerably to the computational intensity (we show these results in Table 7). Since the quantile regressions alone already take more than a day to run, we choose not to report panel-robust standard errors.