Netspar

Network for Studies on Pensions, Aging and Retirement

NETSPAR ACADEMIC SERIES

The Legacy of Interest-Only Mortgages

A Microsimulation Study of the Dutch Mortgage Portfolio

Jan Jakob Lameijer

MSc Thesis 2015-009

The Legacy of Interest-Only Mortgages: A Microsimulation Study of the Dutch Mortgage Portfolio

Master's Thesis Econometrics

Jan Jakob Lameijer

s1904205

January 14, 2015

This research is partly financed by De Nederlandsche Bank. Views expressed are those of the individual author and do not necessarily reflect official positions of De Nederlandsche Bank.

Master's Thesis Econometrics, Operations Research and Actuarial Studies
Specialization: Econometrics
University of Groningen

Supervisors:
Prof. Dr. R.J.M. Alessie (RUG)
Dr. M. Mastrogiacomo (DNB)

Co-assessor:
Prof. Dr. R.H. Koning (RUG)

The Legacy of Interest-Only Mortgages: A Microsimulation Study of the Dutch Mortgage Portfolio

Jan Jakob Lameijer
s1904205

January 14, 2015

Abstract

This thesis examines the risks associated with the large share of interest-only mortgages in the Dutch mortgage portfolio. Using a novel dataset on individual loan characteristics we build a microsimulation model that simulates the mortgage debt up to thirty years in the future. By doing so, we show that voluntary repayments contribute substantially to the repayment of interest-only mortgages and that common risk triggers appear to be favorably distributed across borrowers.

Furthermore, we estimate non-housing wealth for the simulated borrowers and show that most households will not have saved enough to fully repay the mortgage at maturity. Nevertheless, we find that home equity is likely to be positive for these borrowers, such that risks associated to the banking sector are limited. Results are presented for different house price scenarios.

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

Contents

Glossary

AC	Accumulated capital. Some mortgage types accumulate capital in savings accounts pledged to the mortgage. We refer to this amount as the accumulated capital.
Home equity	Property value minus the outstanding mortgage debt. Different value and debt concepts can be used, as explained in Section 3.1.1.
I-O	Interest-Only. A type of mortgage loan where no repayment scheme is attached to the mortgage but only interest is paid. Appendix A provides an overview of all mortgage types in the Netherlands.
LTV	Loan-to-value. It expresses the ratio of the outstanding mortgage debt to the value of the property. Original LTV refers to the LTV at origination date.
Maturity date	The final date of the mortgage contract, typically thirty years after origination.
Mortgagor	Borrower (who lends money to purchase a property).
Mortgage	All outstanding loans secured by the collateral value of real estate property. We differentiate between mortgages and loans, where a borrower typically has a mortgage comprised of multiple loans.
Mortgagee	Lending institution (that lends money to borrowers who want to purchase a property).
NHG	National Mortgage Guarantee [Nationale Hypotheek Garantie]. A public mortgage loan insurance scheme in the Netherlands, where the government acts as guarantor for the mortgage payments.
Origination date	The date at which the mortgage loan was created.
Principal	The amount borrowed. Interest is calculated based on the principal.
RMBS	Residential Mortgage-Backed Securities. Pool of mortgage loans created by lending institutions, which can be purchased by investors. The cash flows are generated from the interest and principal payments on the mortgages.
Starters	First-time home buyers.
Term	The length of the mortgage contract, typically thirty years.
Underwater	A mortgage is underwater when the value of the property is less than the outstanding mortgage debt. In this case the home equity is negative and the LTV exceeds 100%.
Vintage	The amount of years from origination date.

1 Introduction

Recent reforms of the Dutch mortgage market introduced new regulations to reduce the vulnerability of banks and households associated with the large mortgage portfolio. For example, non-amortizing mortgage loans are no longer eligible for the mortgage interest tax-deduction, encouraging households to repay their debt. However, these regulations mainly apply for new mortgages, such that some of the risks regarding current homeowners remain. Still more than 50% of the Dutch mortgage portfolio consists of interest-only loans, where no repayments on the principal are requested (CPB, 2014). Especially in the long run these loans could impose considerable risks, when the maximum amount of years that homeowners are entitled to the tax benefits will be exceeded. This would imply, for instance, an increase of net monthly costs to mortgagors and a substantial amount of debt left in the banks' books beyond maturity. How much do we need to worry about these interest-only mortgages? Will households with an interest-only mortgage be able to pay-off their debt at maturity? These questions are still on the periphery of current policy debates and this thesis aims to contribute to this debate.

Moreover, according to the Dutch Central Bank (DNB [De Nederlandsche Bank]), the Dutch mortgage market has a Janus-faced profile (DNB, 2014). On the one hand the portfolio is characterized by a large number of underwater mortgages, high loan-to-value (LTV) ratios and a large share of interest-only loans. On the other hand, defaults and foreclosures are among the lowest in the world. This thesis attempts to resolve this apparent paradox, where we update and extend results found in Mastrogiacomo and van der Molen (2015).

Specifically, using the loan level data (LLD) gathered by DNB we are able to disaggregate housing wealth, shedding light on the accumulated savings and assets pledged as collateral for the mortgage. The clear 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. Moreover, the data covers an impressive 80% of the entire mortgage portfolio and contains information on a large number of attributes, such that we are able to evaluate how risks are distributed across households. To summarize, 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 loans in the Dutch mortgage portfolio?

In order to provide a more specific answer to the rather broad question above we formulate three sub-questions:

1. *How are interest-only loans distributed across the Dutch homeowners?*

2. *How much of the current mortgage debt will be redeemed in the coming thirty years?*

3. *Will households with interest-only loans have saved enough to pay off their mortgage at maturity?*

To answer these questions we build a microsimulation model that simulates the mortgage debt at borrower level up to thirty years in the future, where we use 2013 as the base year. We neglect the inflow of new mortgagors, as for them annuity contracts are the rule, repaying the mortgage in its entirety. 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.

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. Mortgages that are currently underwater are typically amortizing mortgages (at least partially). In effect, 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. Moreover, 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.

Another contribution of this thesis is that it relates mortgage debt to non-housing wealth. Non-housing wealth is not observed in the LLD, while mortgage information is poorly reported in most datasets that include wealth. This implies that, for instance, the financial equities pledged to the mortgage and overall financial wealth are hardly ever studied at the same time. 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. Subsequently, we use this model to impute net savings in the LLD. By doing so, we find that many borrowers 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 substantial remaining debt (roughly 50 000 euro on average). Moreover, as retirement is likely to occur soon afterwards, these borrowers may be confronted with a drop in income as well. Therefore, the debt service ratio of these households may worsen due to rising costs and falling incomes. Downsizing might be a good option for these borrowers, given that the home equity at maturity is likely to be positive for almost all borrowers. The associated risks to the banking sector should be limited.

The topic of this thesis also closely relates to several other issues currently discussed. Think for instance about the proposal to further reduce LTV caps below 100% after 2018, the possibility to increase mortgage loans risk weights, or the further selection of high quality mortgages into securitized pools. While we do not specifically aim to contribute to any of these discussions, our results show that LTV reductions take place also thanks to voluntary repayments. We show also that the most common risk triggers appear to positively sorted, which should possibly please the supporters of low risk weights.

The remainder of this thesis 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, followed by a brief discussion in Section 6. Finally, Section 7 concludes this thesis.

2 Characteristics of the Dutch mortgage market

The Dutch mortgage market is characterized by high LTV ratios and a large percentage of interest-only mortgages compared to international standards DNB (2014). However, each national mortgage market has its specific features, such that international comparison alone could be uninformative about the associated risks. To provide some background, we will elaborate on the latest developments in the Dutch mortgage market, the risks involved with the mortgage portfolio and the rationale for recent policy regulations.

2.1 Main trends

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. Secondly, banks allowed borrowers to increase their mortgages due to the expected increase in collateral value and, in turn, households used their extended capacity to accumulate debt mainly for housing purposes.[1] 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. Moreover, a variety of new complex loan products were introduced which enabled borrowers to defer repayment of the principle and therefore ex-

[1] 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.

ploiting the mortgage tax-deductibility as much as possible.[2] Lastly, 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 up to a maximum amount (currently 265 000 euro).

2.2 Main risks

Eventually, the bursting of the housing bubble in 2008 revealed the vulnerabilities of the Dutch economy. 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).

The resulting risks are mainly borne by households, banks and the government. First, households with an underwater mortgage are left with a residual debt when selling the house, reducing their mobility. As a result, one could be forced to reject a new job far from home. On top of that, households have experienced a negative wealth effect due to the decrease in house prices, which has a negative impact on household consumption. Especially the poorer and more highly leveraged households contribute to this impact, as their marginal propensity to consume out of housing wealth is substantially higher (Campbell and Cocco (2007); Bostic et al. (2009)). Secondly, 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 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.

2.3 Main regulations

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

[2]see Appendix A for an overview of the different mortgage types in the Netherlands.

fully amortizing mortgages are eligible for the interest deduction. Moreover, 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. Currently, this LTV cap is set to 103%, which will gradually reduce to 100% in 2018. Also, the Financial Stability Committee (FSC) is currently discussing about lowering the limit even further to 90%. Imposing such a limit on the LTV will dampen the procyclical movements in house prices and therefore enhance financial stability. An LTV restriction might generate negative side-effects as well, such as a decrease in house prices due to a lower demand for owner-occupied housing (at least in the short run). 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.

3 Data

The morel feature of this study is the use of the loan level data (LLD) collected by DNB. This section will therefore extensively describe the limitations, advantages and quality of this novel dataset and present some insightful preliminary statistics. Furthermore, a brief description of the Income Panel Study (IPO [Inkomens Panel Onderzoek]) gathered by the CBS is presented, together with some descriptives on non-housing wealth.

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 of the ECB, it is to some extent not designed for analytical purposes. Some important variables are missing or need to be manipulated for our analyses. However, given the granularity and detailing of the LLD, less assumptions are needed relative to other datasets currently available. Mastrogiacomo and van der Molen (2015) extensively

[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)

	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 I-O	44.98%	-	44.03%	-	43.08	-
Excluding I-O	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	

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.

describe some limitations and advantages of the LLD and give interesting new views on the Dutch mortgage portfolio. This thesis updates and extends part of their results.

The template is reported quarterly. 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 interest-only (indicated as I-O), 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 Section 3.1.2 we will also present the debt-weighted shares per loan type, which provides an even more complete picture.

A second advantage of the LLD is the coverage. Using information on the number of mortgages in the Netherlands provided by Statistics Netherlands, we estimate that the LLD covers approximately 80% of the total population, as shown in Table 1. The non-reported mortgages are those of small banks, pension funds, foreign institutions and from institutions not able to report their whole portfolio (and only reporting their securitized pool). In total, eleven institutions have reported (part of) their mortgage portfolio. However, as shown in the table, not all institutions have reported their portfolio in all waves, thereby limiting the potential for panel analyses. Fortunately, about 90% of the mortgages in our sample come from the main banks which have consistently reported their portfolio.

Thirdly, 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. Other relevant information included in the dataset are the outstanding balances – current and original –, origination and maturity date, current mortgage interest rate, mortgage type[4], most recent property valuation amount, year of birth of the mortgagor and income at loan origination.[5]

The data also presents some challenges. As retrospective information is included, it is possible to approximate past developments of the mortgage portfolio. However, a few remarks should be made. First, we observe this information only for a selected group, namely those loans that still survive in the portfolio. When a loan is fully repaid it disappears from the dataset. Second, whenever a loan contract is renegotiated, for instance when moving into a new dwelling or setting a new interest rate, the information about the loan at origination is updated and older information is lost. This translates, for example, into more than 50% of the loans having a vintage lower than 9 years and less than 50% of the borrowers being younger than 41 years at first loan origination. Appendix B elaborates on this and shows how we can still accurately determine aggregate figures such as the historic development of the LTV for starters (i.e. first-time home buyers).

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.[6] 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. The exact calculations and underlying assumptions are described in Section 4. In the summary statistics that follow, we should always keep in mind that the net and gross mortgage debt are approximations.

[4]Mortgage type identification is not always clear, since the mortgage type categories as presented in the data are rather general. For more details, see Appendix A.

[5]Some fields in the RMBS template are indicated as optional, among which the birth year of the mortgagor and income at loan origination. Most banks consistently report the birth year of the mortgagor. Income, however, is only reported for 50% of the borrowers.

[6]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.

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

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

3.1.1 Definitions and concepts

This section presents a detailed discussion on some relevant concepts in our analyses, where the LLD does not always allow for consistent definitions. Moreover, concept definitions may differ slightly between the LLD and IPO. We start with the definition of LTV, which has already been briefly introduced. In its most general form, the LTV is defined as

$$\text{LTV} = \frac{\text{mortgage debt}}{\text{property value}} \times 100\%.$$

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 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])[7] determined by the taxing authority and the liquidation value.[8]

Unfortunately, the LLD does not necessarily allow for a consistent definition of the LTV, 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. For instance, drive-by and desktop

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

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

valuations are typically used when there is a lot of equity in the property, explaining the large average property value indicated in the table. From now on we will use the property value as reported in the LLD, keeping this possible inconsistency in mind. Section 3.2 provides a useful comparison with the IPO data, where the WOZ-value is reported for all observations. A final remark regarding the valuation of the property is that only the most recent valuation amount is reported in the LLD, where a valuation occurs only once in a few years. To this end we use the Dutch house price index from Statistics Netherlands to approximate the current value of the dwelling.

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, as will be discussed in Section 4.1. 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. For example, when part of the mortgage is used to finance non-housing consumption, this is not included in the fiscal debt concept, in contrast to the mortgage debt in the LLD. Finally, 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.

3.1.2 Descriptive statistics

This subsection presents some descriptive statistics based on the 2013 Q4 wave, as this wave provides the most recent picture of the Dutch mortgage portfolio and will be the base year of our simulation. 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 in Section 4.1) yields an estimate of the net mortgage debt of 609 billion euro.

The pie chart on the left 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. The difference between the 50% indicated in (CPB, 2014) can be attributed to the difference in net and gross mortgage debt.

The risks associated with this large portion of interest-only loans might be better explained by showing how these loans are distributed across households. To this end we calculate the debt-weighted share of interest-only loans per borrower and round that number to the nearest even decimal point. By doing so, borrowers are divided in six interest-only categories as presented in the right pie chart in Figure 1. From this figure we observe that 32% of the net mortgage debt comes from borrowers having a 100% interest-only mortgage,

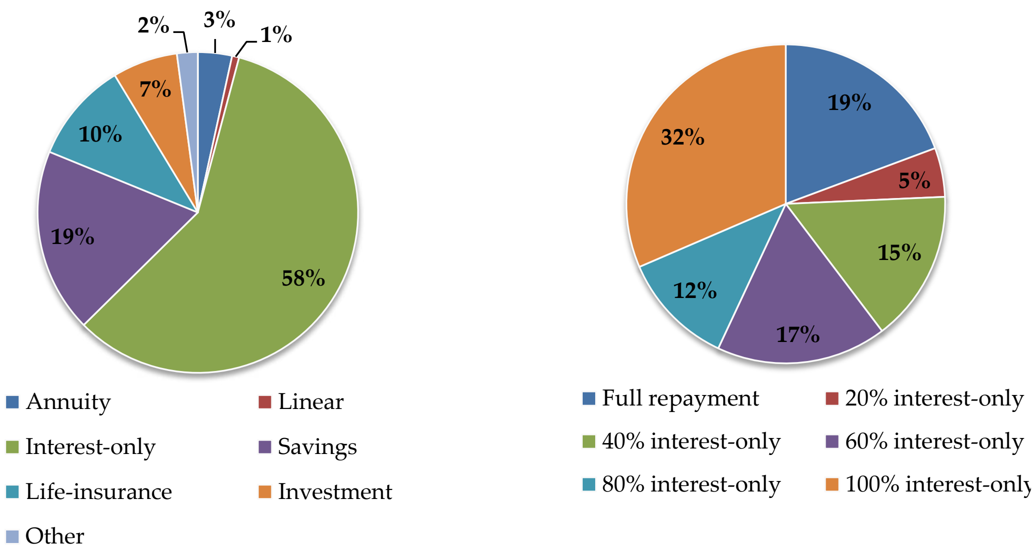

Figure 1: Debt-weighted share of each mortgage type (left) and the debt-weighted share of borrowers per interest-only category (right), both based on net mortgage debt in 2013 Q4.

which is even less than the 35% from Table 1. This interesting result therefore again shows that the large part of interest-only mortgages are often combined with other amortizing mortgages, which nuances the view that most households never amortize.

Descriptive statistics for the relevant variables in our study are presented in Table 3, where descriptives are given per interest-only category.[9] 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. A few interesting insights are obtained from this table. First, we observe that borrowers having a large share of interest-only loans are typically older borrowers. These borrowers have often bought their first property a long time ago before the sharp price increase in the 1990s, after which they experienced a large growth in the value of their property. To this end, we observe that borrowers in this category also have a relatively high property value and a small mortgage debt, translating into a low LTV. An interest-only loan was often used to cash out home equity, which could partly explain the over-representation of older borrowers in the highest interest-only category. Also, older borrowers could have simply repaid other amortizing loans already. Only 5% of the borrowers in this category have a mortgage that is underwater, which again nuances the risks associated with the large part of the interest-only mortgages. Also, it is not hard to believe that older borrowers have more financial assets, which we will investigate in this thesis as well.

We furthermore observe that this relationship between interest-only share and LTV is not linear. On average, the youngest borrowers fall in the 40% interest-only category, where the

[9]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.

	0% I-O			20% I-O			40% I-O		
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 3: Descriptives LLD 2013 Q4 on borrower-level per I-O category

average LTV is no less than 93% and where 54% of the mortgages are underwater. However, these borrowers do contractually amortize on more than half of their mortgage debt. The simulation described in Section 4.3 attempts to show how this affects the development of the LTV for the upcoming years. Also, we find that a large share of underwater mortgages is often accompanied with a large share of mortgages that are NHG-guaranteed. This indicates that the risks associated with underwater mortgages are to a significant extent borne by the government, which again limits the risks to borrowers and lending institutions.

Voluntary repayments are not directly observed, but can be retrieved by taking the difference in mortgage balance between 2012 Q4 and 2013 Q4, where we correct for contractual mortgage repayments. Small differences (between -50 and 50 euro) are treated as zero, as they mainly occur due to the fact that the estimated contractual repayments are indeed approximations. Borrowers can increase their mortgage as well, for example to finance renovations or to cash home equity. We assume that a mortgage take-up and voluntary repayments are mutually exclusive mechanisms determined by other factors, such that we simply cannot observe voluntary repayments for borrowers which have increased their mortgage. Altogether, we lose 473 980 (710 714) borrowers in 2013 Q4 (2012 Q4), yielding a restricted sample of 1 901 566 borrowers for which the voluntary repayments are observed. The removed observations do not constitute a random sample of all observations, as selection is based on observables. For example, borrowers increasing their mortgage might be more risky. Also, borrowers only observed in 2013 Q4 are likely to be first-time home buy-

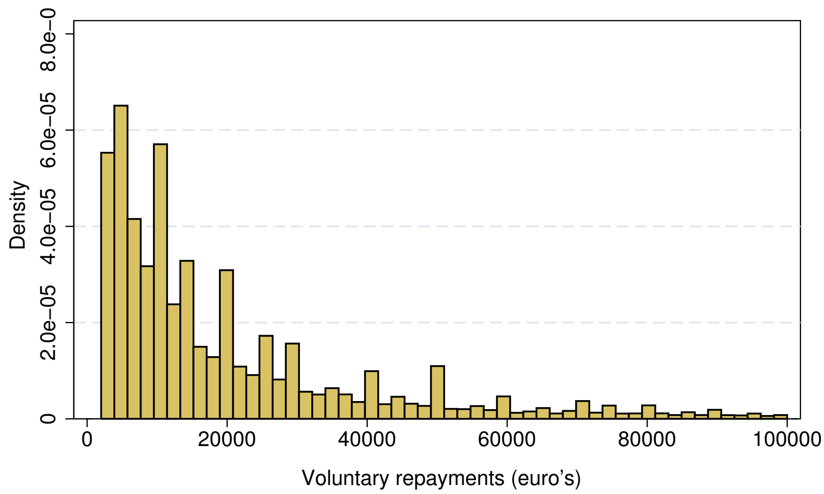

Figure 2: Distribution of the voluntary repayments in 2013 (truncated at 100 000 euro)

ers, typically having a higher LTV. Nevertheless, to estimate future voluntary repayments we are forced to base our model on this restricted sample, thereby implicitly assuming the data generating process for voluntary repayments is similar for the non-observed group.

By taking the yearly difference we remove all seasonal components. However, given the limited number of waves in the LLD we can only observe 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.

Moreover, considering the administrative costs of processing, most lending institutions have set a lower limit for voluntary repayments. This minimum amount differs across and within institution, but almost all lending institutions apply a lower limit less than 2 000 euro per year. To this end we consistently treat all voluntary repayments 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).

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 on aggregate level, representing roughly 2% of the net mortgage debt. A histogram of the resulting (non-zero) voluntary repayments is provided in Figure 2. The peaks indicate that round numbers are more popular amounts to voluntarily repay, as expected. As is common in expenditure data, the distribution of the positive voluntary repayments is highly right-skewed (skewness is 5.76) and has considerable non-normal kurtosis (kurtosis is 74.51). This might be an indication that voluntary repayments are best modeled using the natural logarithm (but could also be due to skewed regressors).

	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 4: Descriptives IPO 2005, 2008 and 2011

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. The IPO dataset is an administrative panel dataset containing a representative sample of Dutch households. The sampling method is based on social security number, after which the selected persons are followed over time. Each year, the sample is extended by newborns and immigrants. A selected person may exit the panel by emigration or by death. The data is gathered ultimo December of each year and contains observations on demographic characteristics, income and wealth for each household member of the selected persons. See CBS (2014a) and CBS (2014b) for details on all observed variables.

In total, the dataset consists of 1 852 323 observations, containing information on 112 942 unique households. However, we only select the household heads that own a property financed by a mortgage. Also, 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.

As being said, the missing information 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. To this end, 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 in Section 4.2, they alert us that difficulties may arise when modeling net savings.

Moreover, Figure 3 presents age and cohort patterns of the net savings, where we use five-year birth cohorts. Birth years 1923 - 1927 are for the oldest cohort and 1988 - 1992

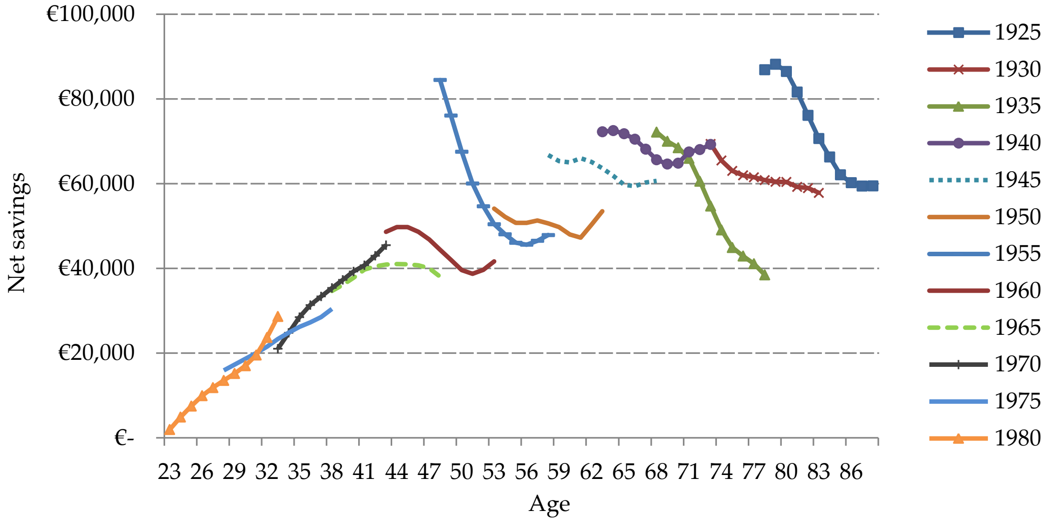

Figure 3: Net savings by birth cohort using nonparametric locally weighted regression (LOWESS) with a bandwidth of 0.8. Labels correspond to the middle year of each cohort.

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. Especially for older cohorts, where the vertical differences are larger, there appears to be cohort or time effects.

Figure 4 compares the distribution of the property value as observed in both the 2011 wave of the IPO and the 2013 Q4 wave of the LLD, where the values are indexed to 2011 for the latter dataset. Reassuringly, the distributions are very similar. The distribution in the LLD is a bit more shifted to the right, which can be attributed to the different value concepts used in the LLD as explained in Section 3.1.1. Also, a comparison of the distribution of the gross mortgage debt is provided in Figure 5. The distributions are again very similar, where the small difference can be attributed to the different debt concepts and the gap of two years. Hence, although we are not able to perfectly align our value and debt definitions, these results indicate that this should not be a big problem.

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]).

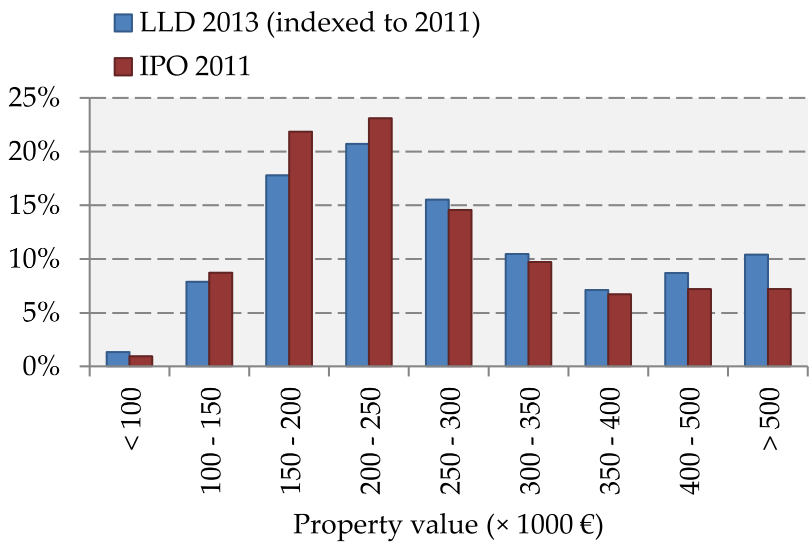

Figure 4: Distribution of the property value in both LLD and IPO

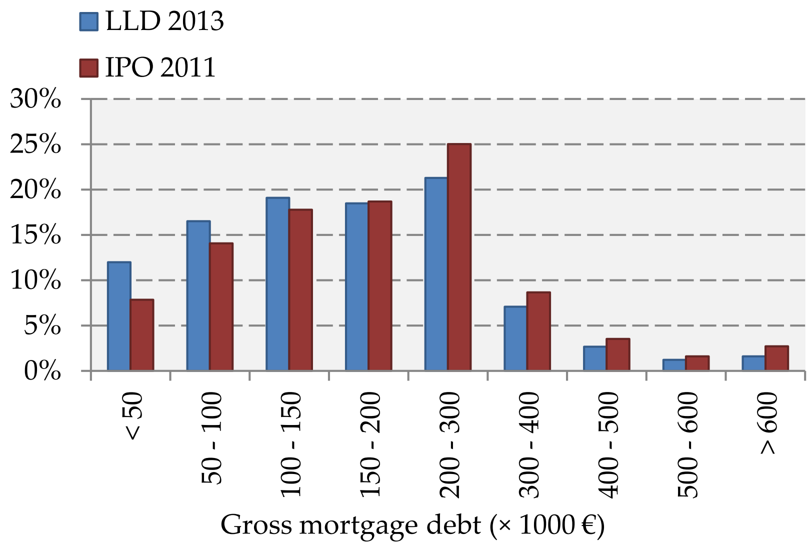

Figure 5: Distribution of the gross mortgage debt in both LLD and IPO

4 Methodology

This section describes the design of the microsimulation model. First, we explain how we model the different components that determine the net mortgage debt of each borrower. Next, we show how the IPO dataset is used to estimate a model for net household savings. Finally, we explain how these models are combined in the overall simulation method, where we run a number of different scenarios to test the sensitivity of the results to the modeling assumptions.

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. Recall that we only focus on the future contribution to household debt of currently existing mortgages.

4.1.1 Contractual mortgage repayments

The only two mortgages that contractually amortize are the annuity and linear mortgage. In order to calculate the yearly contractual repayment for these two mortgage types we define the following parameters:

r	=	annual mortgage interest rate,
A	=	annuity in euro,
d	=	mortgage term in years,
t_0	=	mortgage origination year,
R_t	=	mortgage repayment at time $t \in \{t_0+1, \ldots, t_0+d\}$,
D_t	=	mortgage debt at time $t \in \{t_0, \ldots, t_0+d\}$.

Without loss of generality we assume that $t_0 = 0$. For the linear mortgage, the repayment part is constant over time and the mortgage is fully repaid at maturity. The yearly contractual repayment therefore equals D_0/d. To calculate the contractual repayment scheme for the annuity mortgage we assume that the interest rate remains constant throughout the whole period, as we have no information on possible interest rate changes. Now, by definition, A is constant over time as well and consists of the interest paid on the mortgage and the mortgage repayment part, such that we have

$$A = R_t + r \cdot D_{t-1},\ t = 1, \ldots, d.$$

Substituting this into

$$D_t = D_{t-1} - R_t,$$

yields

$$D_t = (1+r)D_{t-1} - A. \tag{1}$$

Recursively solving for D_d gives

$$\begin{aligned} D_d &= (1+r)D_{d-1} - A \\ &= (1+r)^2 D_{d-2} - (1 + (1+r))A \\ &\vdots \\ &= (1+r)^d D_0 - \sum_{i=0}^{d-1} (1+r)^i A. \end{aligned}$$

Finally, using $D_d = 0$ we can solve the above equation for A to get

$$A = \frac{(1+r)^d D_0}{\sum_{i=0}^{d-1}(1+r)^i} = \frac{-r(1+r)^d D_0}{1-(1+r)^d},$$

where the right-hand side of the equation is known. Now, using A and D_0 in (1) we can determine the mortgage debt of an annuity mortgage in each time period.

4.1.2 Capital accumulation in savings accounts pledged to the mortgage

As discussed, there are multiple mortgage types that accumulate capital pledged to the mortgage. First, we consider only the savings mortgage loans, where we know that the accumulated capital at maturity must be enough to fully repay the mortgage. We use the same parameter definitions as in the previous section. Additionally, we define AC_t as the total accumulated capital at time t and S as the savings premium. Again, we have to assume that the interest rate remains constant over the whole period. Moreover, we assume that the mortgage interest rate is the same as the interest rate on the savings account. Unfortunately, we are not able to determine the prepayments in the savings account, such that we have to assume there is no capital in the savings account at $t = 0$. This leads to an underestimation of the accumulated capital and an overestimation of the savings premium for some loans. Given these assumptions, we have

$$AC_t = \sum_{i=1}^{t} S(1+r)^i,$$

for $t \in \{1, \ldots, d\}$. Here, we can determine S based on the requirement that the accumulated capital will be enough to repay the mortgage at maturity, i.e.

$$\begin{aligned} D_0 &= AC_d \\ &= \sum_{i=1}^{d} S(1+r)^i \\ &= S\left(\frac{1+r-(1+r)^{d+1}}{-r}\right). \end{aligned}$$

Solving for S yields

$$S = \frac{-r \cdot D_0}{1+r-(1+r)^{d+1}},$$

where the right-hand side of the equation is known, such that we are able to determine the accumulated capital in all time periods.

For life insurance mortgages, part of the return depends on the investment returns of the insurer, such that it is uncertain whether you can repay the whole mortgage at maturity. However, a minimum return is often guaranteed, and most of the time the accumulated capital will be close to the total mortgage debt. To this end we will treat life insurance mortgages similarly to savings mortgages, thereby implicitly assuming that the accumulated capital at maturity will be enough to repay the mortgage.

For investment mortgages, however, the capital accumulated at maturity can differ widely from the principal amount and is often close to zero. Moreover, we have no information on the structure of an investment mortgage. For example, one could invest a lump sum at the beginning of the mortgage period, or contribute periodically to the investment account. To this end, we simply assume investment mortgages do not accumulate capital in a separate account. This might seem extreme, but is often the case in practice due to very disappointing investment returns and large insurance costs. From a financial stability perspective, this is perhaps the most relevant assumption. To test the sensitivity to this assumption, we also consider the scenario where investment mortgages are treated similar to savings mortgages.

Finally, interest-only mortgages do not accumulate capital and we treat all "other" mortgages, as defined in the data, similar to an investment mortgage.

4.1.3 Voluntary repayments

Recall from Section 3.1.2 that we only have cross-sectional data available to model voluntary repayments, where we use the observations in 2012 Q4 to explain the voluntary repayments over the coming year. Here, we have $N = 1\,901\,566$ borrowers for which the voluntary repayments are actually observed. Given the limited data availability we have to assume independence over time, but more sophisticated models can be considered once more waves will be available in the future. For notational convenience we therefore drop the subscript t when estimating the model. Let y_i denote the voluntary repayments for borrower $i = 1, 2, \ldots, N$. 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):

$$y_i^* = \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i, \ i = 1, 2, \ldots, N, \tag{2}$$

where y_i^* is the latent response variable underlying y_i and $\boldsymbol{x}_i$ is an $(k+1) \times 1$ vector of a constant and k explanatory variables for which $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{k+1})'$ are the corresponding coefficients. Moreover, $\varepsilon_i \sim NID(0, \sigma^2)$ and independent of $\boldsymbol{x}_i$. 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 $L = 2\,000$. 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. To this end we simply 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

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

where the last step follows since the distribution of ε_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). Specifically, Appendix C explains how we can use a simple auxiliary regression to perform an asymptotically equivalent Lagrange Multiplier test for heteroskedasticity, where the test for normality is constructed by the same kind of reasoning. If the error term of the latent equation is heteroskedastic or not normally distributed, Tobit maximum likelihood estimates are inconsistent.

Moreover, the probit and Tobit should yield similar parameter estimates, as they are based on the same latent model. Notice, however, that σ 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 $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_{k+1})' = ((\beta_1 - L)/\sigma, \beta_2/\sigma, \beta_3/\sigma, \ldots, \beta_{k+1}/\sigma)'$. 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 $\Pr\left(y_i > 0|\boldsymbol{x}_i\right)$ with the ME from the probit model. Let x_{ij} denote the jth component of $\boldsymbol{x}_i$. Now, the ME of change in regressor x_{ij} on $\Pr\left(y_i > 0|\boldsymbol{x}_i\right)$ is given by

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

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

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

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. To this end, 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

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

where we should note that $\boldsymbol{\beta}$, ε_i and σ^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 $\ln y_i$.[10]

The Tobit model has some restrictive implication, e.g. the ME of x_{ij} on $\Pr(y_i > 0|\boldsymbol{x}_i)$ and $\mathrm{E}(y_i|\boldsymbol{x}_i, y_i > 0)$ always have the same sign. By relaxing these assumptions we might obtain a better fit. To this end 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:

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

where $I(.)$ is the indicator function, $v_i|\boldsymbol{x}_i \sim NID(0, 1)$ and $u_i|\boldsymbol{x}_i \sim NID(0, \sigma^2)$ 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 σ^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 (as explained in Section 4.3).

[10] 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.

4.2 Non-housing wealth

Recall that we use panel data to model net household savings, where we observe $N = 63\,791$ borrowers over $T = 7$ years. The panel is unbalanced, such that the total number of observations $S = 341\,118 < NT$. Now, let y_{it} denote the net savings for borrower i at time t. Net savings might be very difficult to model, given its unique nature. For example, net savings might be influenced to a large extent by luck (think about inheritance or lottery winnings). Moreover, 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:

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

where c_i is an unobserved individual effect, u_{it} is an error term and $\boldsymbol{x}_{it}$ 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 $\mathrm{E}(u_{it}|\boldsymbol{x_{it}}, c_i) = 0$. 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 $\mathrm{E}(c_i|\boldsymbol{x}_i) = 0$, where $\boldsymbol{x}_i = (\boldsymbol{x}_{i1}', \ldots \boldsymbol{x}_{iT}')'$. If we make the fixed effect assumption instead, i.e. $\mathrm{E}(c_i|\boldsymbol{x}_i) \neq 0$, 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 $\boldsymbol{x}_{ij}$. 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, finally, fifteen cohort dummies.

Now, let $v_{it} = c_i + u_{it}$ such that (7) can be rewritten as $y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + v_{it}$. The assumptions on u_{it} and c_i imply that $\mathrm{E}(v_{it}|\boldsymbol{x}_{ij}) = 0$, such that the conditional expectation of y_{it} is given by $\mathrm{E}(y_{it}|\boldsymbol{x_{it}}) = \boldsymbol{x}_{it}'\boldsymbol{\beta}$. Moreover, $\mathrm{E}(v_{it}|\boldsymbol{x}_{ij}) = 0$ is sufficient to prove that β 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 thesis 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 $N_b = 10\,000$ 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 $\boldsymbol{x}_{it}$, yielding B estimates of $\boldsymbol{\beta}$ denoted by $\widehat{\boldsymbol{\beta}}_b, b = 1, \ldots, B$. 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

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

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 qth conditional quantile of the distribution of $y_{it}^{\bullet}$ by $Q_q(y_{it}^{\bullet}|\boldsymbol{x}_{it})$, where we assume $Q_q(y_{it}^{\bullet}|\boldsymbol{x}_{it})$ is linear in $\boldsymbol{x}_{it}$, i.e. $Q_q(y_{it}^{\bullet}|\boldsymbol{x}_{it}) = \boldsymbol{x}_{it}'\boldsymbol{\beta}_q$. 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:

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

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 $Q_q(y_{it}^{\bullet}|\boldsymbol{x}_{it})$, we simply transform this estimate using the hyperbolic sine function to get an estimate for $Q_q(y_{it}|\boldsymbol{x}_{it})$. Again, the bootstrap method should be used to obtain panel-robust standard errors, which adds considerably to the computational intensity. Since the quantile regressions alone already take more than a day to run, we choose not to report panel-robust standard errors.

4.3 Simulation method

We start our simulation in 2013 using the borrowers from the LLD observed in 2013 Q4. 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 provided in Section 5.1.1, this will be done according the Cragg log-normal hurdle presented in equation (6). 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 $\widehat{\sigma}^2$. Here, $\widehat{\sigma}^2$ is the estimated variance of u_{it} from equation (6). Finally, the exponential function is used to transform the repayment amount back to levels.

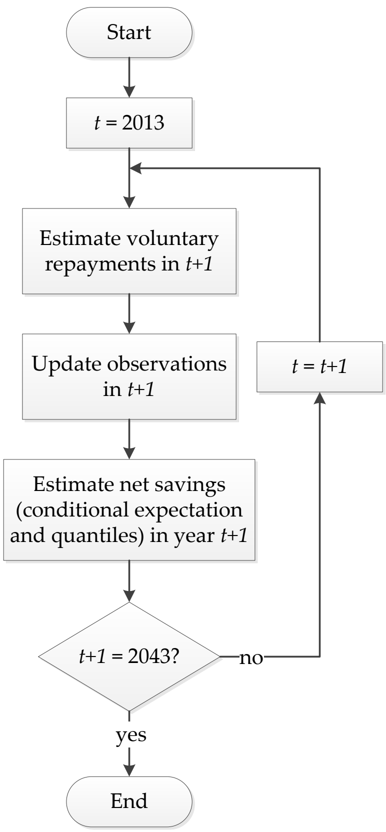

Figure 6: Flowchart of the microsimulation per borrower

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/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. Naturally, we assume the borrower will not repay more than he has debt. The contractual mortgage repayment and capital accumulation are calculated as described in Section 4.1. 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. Moreover, using the resulting updated variables and the estimated models from Section 4.2 we can estimate the conditional mean and conditional quantiles of the distribution of net savings for each borrower in all simulated years. In our analysis of the results we will mainly focus on the distribution of net savings at maturity.

5 Results

First, we present the estimation results for the regression models of both voluntary repayments and net savings, where we explain which models to use in the simulation. Subsequently, a series of interesting results from the overall microsimulation are provided.

5.1 Estimation results

5.1.1 Voluntary repayments

The first two columns in Table 5 present 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. To illustrate the interpretation of the ME we take age as an example: on average, the probability that a borrower makes a voluntary repayment decreases by roughly 0.001 if age increase by one year, holding other factors constant. 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. Moreover, the LM tests strongly reject the hypothesis of homoskedasticity and normally distributed error terms. Again, the rejection might be

	Probit (Part I)		Tobit (in levels)		Tobit (in logs)		Two-Part (Part II)
Regressors	Coef	ME	Coef	ME ($\Pr(y_i > 0 \mid \boldsymbol{x}_i)$)	Coef	ME ($\Pr(y_i > 0 \mid \boldsymbol{x}_i)$)	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)
$(\text{age}/10)^2$	-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)
current $\text{LTV}/10^2$	-0.123***	-0.0267***	4709***	0.0148***	-0.102***	-0.00642***	0.914***
	(0.00495)	(0.00107)	(320.0)	(0.00101)	(0.0167)	(0.001.05)	(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		-3 667 000		-2 881 000
$\widehat{\sigma}^2$			65 986				1.048
Two-Part model:							
pseudo R^2							0.010
Log-likelihood							-3 632 000
p-value LM tests:							
heteroskedasticity	0.000						
normality	0.000						

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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

Regressors	Linear Probability Coef	ME	Logit Coef	ME	Probit 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)
$(\text{age}/10)^2$	-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***
	(0.000730)	(0.000730)	(0.00626)	(0.000734)	(0.00337)	(0.000730)
interest rate	0.234***	0.234***	2.143***	0.251***	1.125***	0.244***
	(0.0290)	(0.0290)	(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 $\times$ 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)	(0.000730)	(0.00327)	(0.000709)
current LTV/10^2	-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	1 901 566		1 901 566		1 901 566	
pseudo R^2	0.010		0.010		0.010	
Log-likelihood	-760 934		-750 842		-750 856	

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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

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. As discussed, the misspecification of the Tobit model might be caused due to the skewness and non-normal kurtosis in the distribution of the voluntary repayments. 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 Table 5 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. To this end, and for practical reasons discussed before, we 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 Section 3.1.2 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). The estimation results for these models are provided in Appendix D. However, we should point out that in spite of this additional refinement, we still get a rather poor fit of the model. There is only little variation in the fitted values resulting from the probit model, indicating that the decision to voluntarily repay is modeled predominantly as randomly. It only gives an approximation of the percentage of all borrowers that make a voluntary repayment in a specific year.

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. Basically, 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 vertical differences between the older cohort curves in the figure might be caused by the high sensitivity of the mean to outliers. Inequality in net savings increases with age and inheritance possibly produces more outliers in the distribution of net savings for older borrowers.

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.

	Robust regression	Quantile regression (IHS transformed)		
Regressors	(in levels)	$q = 0.25$	$q = 0.50$	$q = 0.75$
age	526.08**	0.0947***	0.0750***	0.0802***
	(265.13)	(0.00979)	(0.00584)	(0.00525)
$(\text{age}/10)^2$	-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)	(0.000249)	(0.0000926)	(0.0000974)
property value/10^3	-2.482	-0.0000586	-0.000161	-0.000161
	(5.806)	(0.000323)	(0.000163)	(0.000109)
interest rate	2698.6	0.526	-0.396	-1.083**
	(10 458)	(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)	(0.000787)	(0.000485)	(0.000405)
I-O share per postcode	-6 310.179***	0.846***	-0.203***	-0.500***
	(4 410)	(0.107)	(0.0685)	(0.0515)
mean house price per postcode/10^3	-0.835*	0.0000109**	0.00000465	-0.000000915
	(2.490)	(0.00000467)	(0.00000380)	(0.00000294)
Average gross mortgage debt/10^2	-3.447***	-0.000334***	-0.000165***	-0.0000919***
	(0.760)	(0.0000233)	(0.0000101)	(0.0000112)
Average property value/10^2	7.275***	0.000351***	0.000413***	0.000444***
	(0.770)	(0.0000335)	(0.0000167)	(0.0000118)
Average interest rate	107 430***	11.40***	9.564***	9.054***
	(28 208)	(1.029)	(0.431)	(0.718)
Birth cohorts	Yes	Yes	Yes	Yes
Constant	-24 043***	5.518***	7.507***	8.546***
	(74 470)	(0.308)	(0.288)	(0.156)
N	341 118	341 118	341 118	341 118
R^2		0.0137	0.0343	0.0578
p-value Wald test for joint significance:				
Birth cohorts	0.927	0.000	0.000	0.000
Postcode variables	0.481	0.000	0.000	0.000

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

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

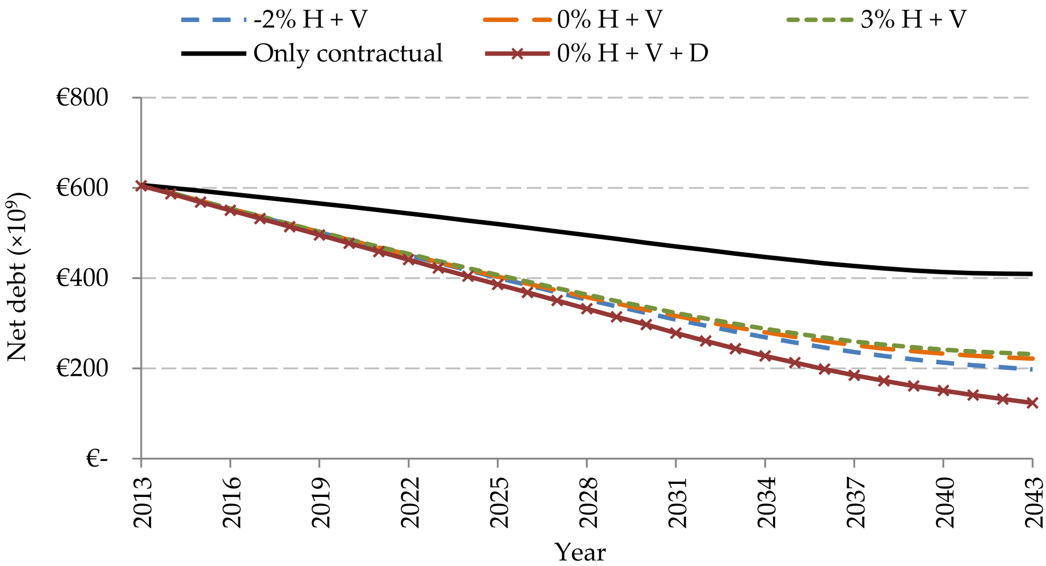

Figure 7: Simulation of the aggregate net mortgage debt for currently existing mortgages in the Netherlands. Different scenarios are considered (H = house price change; V = voluntary repayments; D = mortgage is repaid at death (85 year)).

5.2 Simulation results

The data derived from the microsimulation contains rich information on the mortgage debt and net savings of each borrower for all simulated years. Based on this data we are able to show a series of results yielding interesting insights into the Dutch mortgage portfolio. First, we focus on how much of the current mortgage debt will be redeemed in the coming thirty years. To this end, Figure 7 presents the simulation results of the aggregate net mortgage debt in the Netherlands for different scenarios. Contractual repayments and capital accumulation are a common feature in all 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. 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). The line is almost linear, but where repayments stagnate around 2037. This might imply that most mortgages start to get to maturity around this period, as most mortgages originated around 2007. This is partly explained by the increase in homeownership during this period, but also because the origination and maturity date are overestimated as explained in Section 3.1.

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. Here, we simply assume that borrowers die at age 85 (te average life expectancy of women in the Netherlands), in which case the collateral is used to repay the mortgage. In effect, losses are incurred by the mortgagee when the respective mortgage is underwater (unless the mortgage is combined with an insurance part). We find that 40% of the borrowers in our simulation reaches the age of 85. However, recall from Section 3.1.2 that 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 limited.

Next, we focus on the size of the debt relative to the collateral value, which is of great importance regarding the LGD and the mobility of households. To this end, the development of the average LTV is presented in Figure 8, where different house price scenarios are considered. Moreover, Figure 9 shows the evolution of the share of underwater mortgages. Considering the large share of interest-only loans, it is often thought that the current underwater mortgages remain a persistent problem in the future. We observe that this is partly a misconception. 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 are substantial.

So far we have focused on the characteristics of the mortgage portfolio in all simulated years. From now on we will only concentrate on the mortgage characteristics at maturity, which is an institutionally relevant moment in time as borrowers become no longer eligible to the tax-deductibility. Borrowers who still have a substantial amount of outstanding debt at maturity are therefore confronted with an increase in the debt service ratio, such that default rates might increase as well. Again, 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),

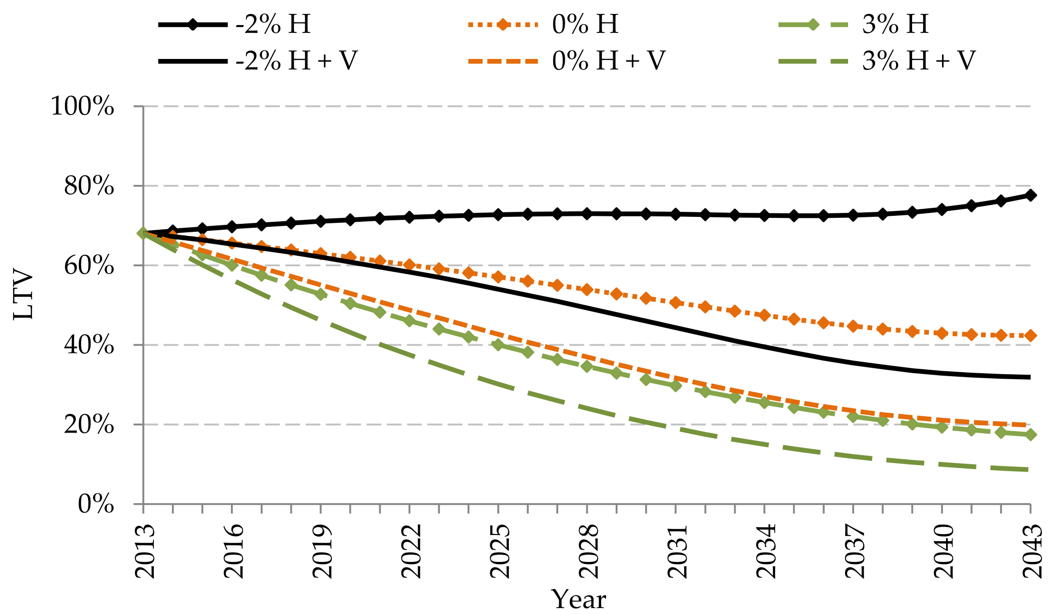

Figure 8: Simulation of the average LTV of the mortgages currently existing in the Netherlands, where different scenarios are considered (H = house price change; V = voluntary repayments).

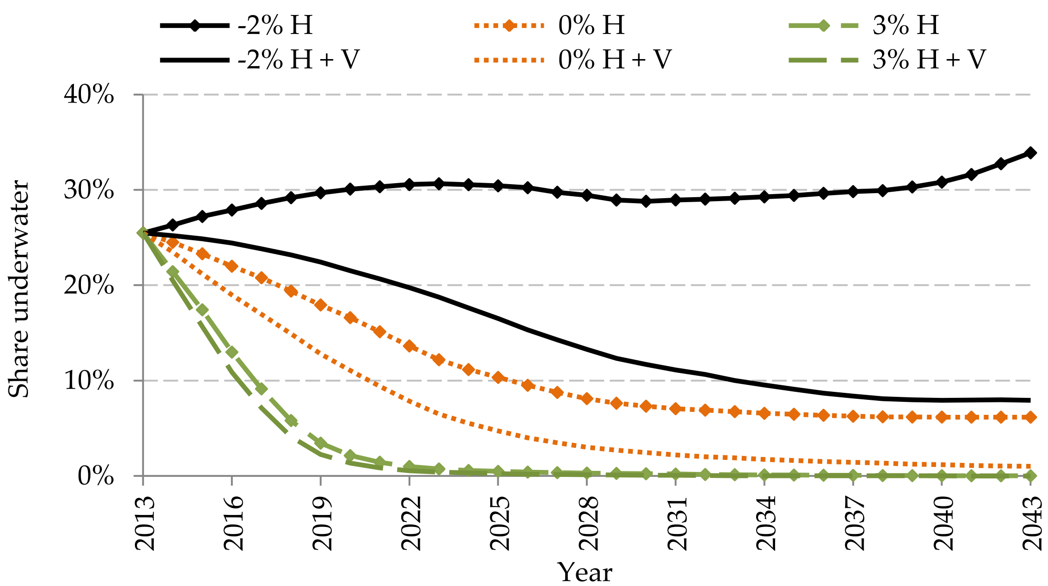

Figure 9: Simulation of the share of underwater mortgages among the mortgages currently existing in the Netherlands, again considering different scenarios (H = house price change; V = voluntary repayments).

we find that only 3% of the mortgages that mature in 2037 are underwater.

Furthermore, 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.

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, thereby avoiding an increase in the debt service ratio. To this end, Figure 14 presents the distribution of both net mortgage debt and net savings per maturity year. 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.

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. Finally, 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 50 000 euro on average. Downsizing might be a good option for these borrowers to (partly) repay the outstanding debt, as we observed that home equity is likely to be positive at maturity.

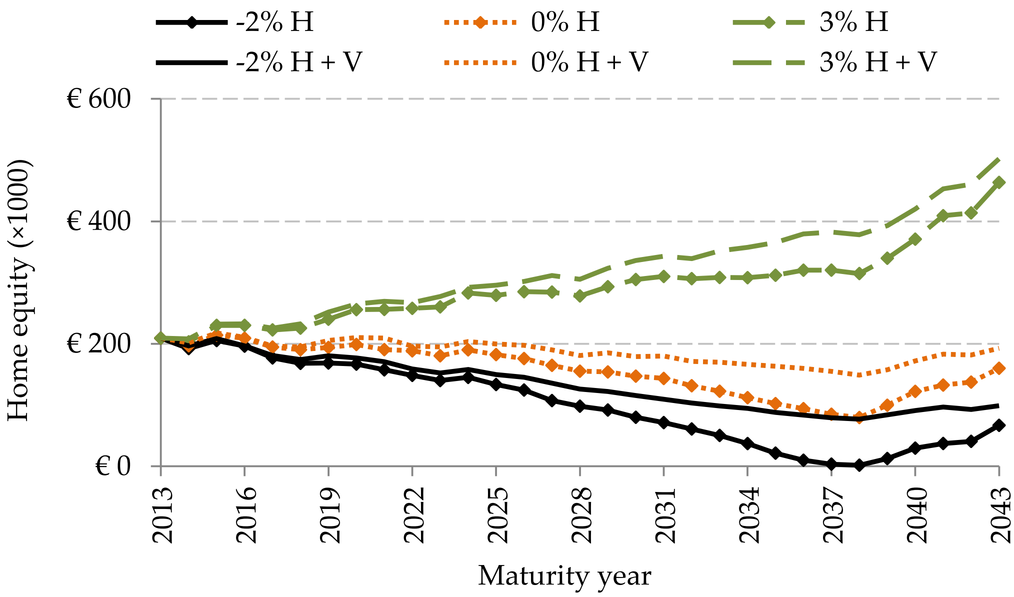

Figure 10: Median home equity per maturity year of the mortgages currently existing in the Netherlands. Different scenarios are considered in the simulation (H = house price change; V = voluntary repayments).

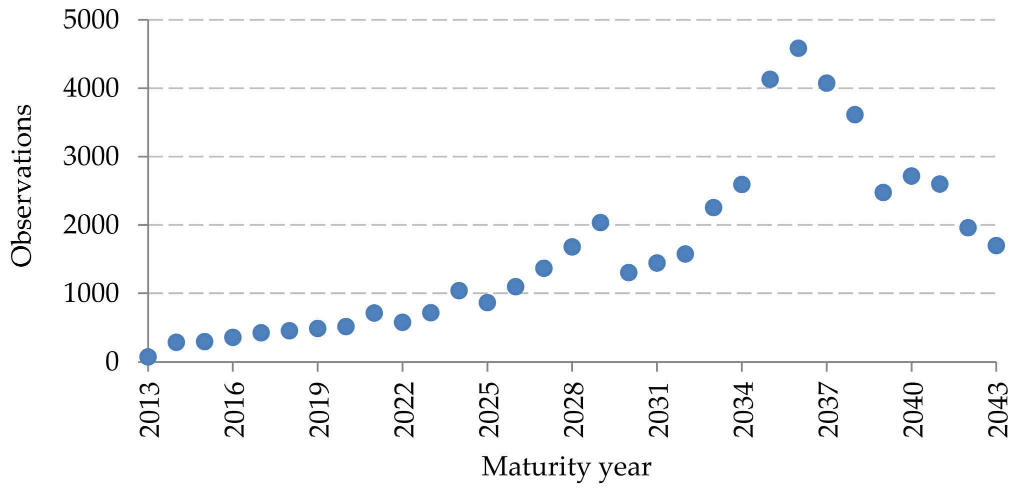

Figure 11: Number of borrowers in the simulation per maturity year of the corresponding mortgage.

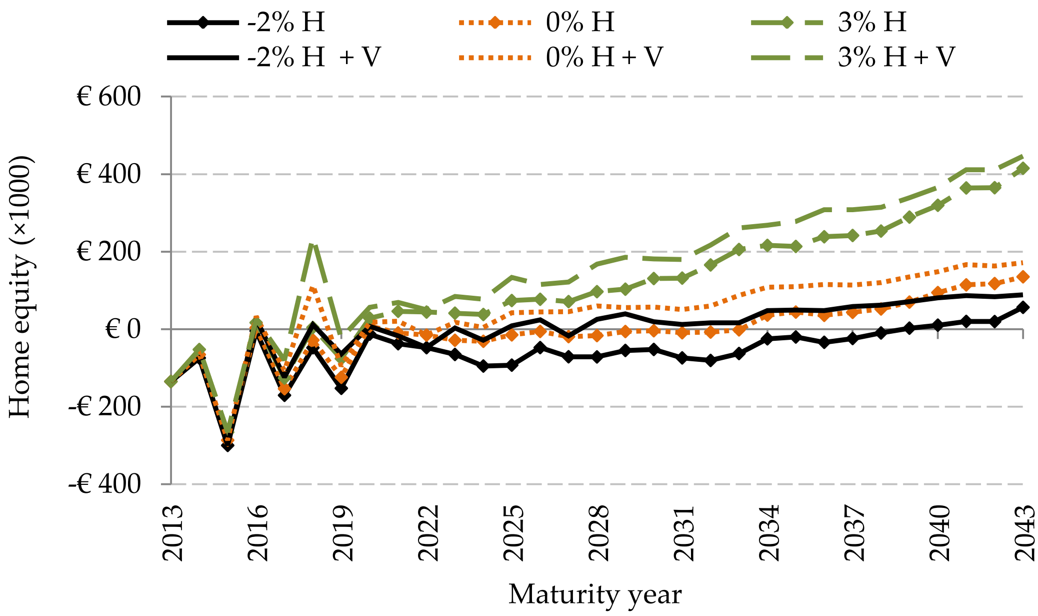

Figure 12: Median home equity per maturity year of the mortgages that are underwater in 2013. Different scenarios are considered (H = house price change; V = voluntary repayments).

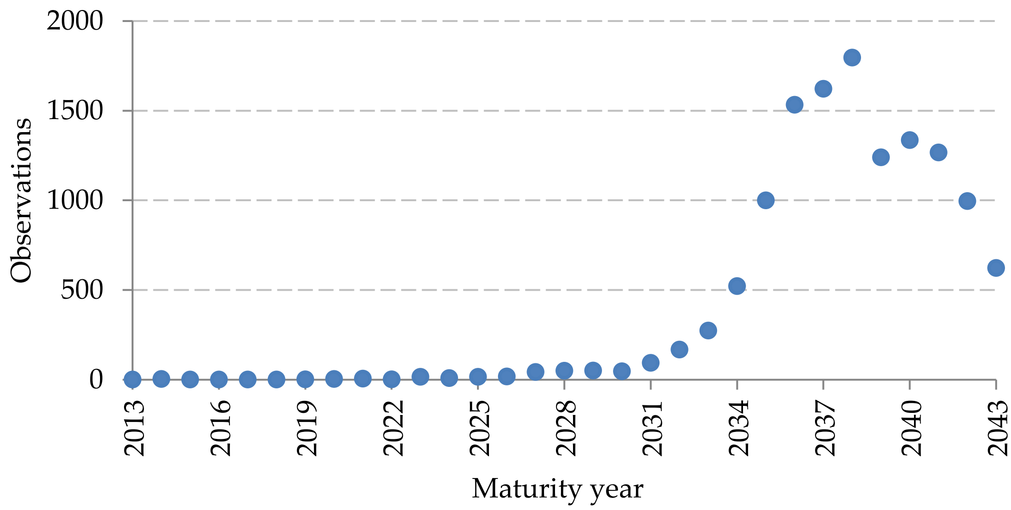

Figure 13: Number of borrowers in the simulation that are underwater in 2013 per maturity year of the corresponding mortgage.

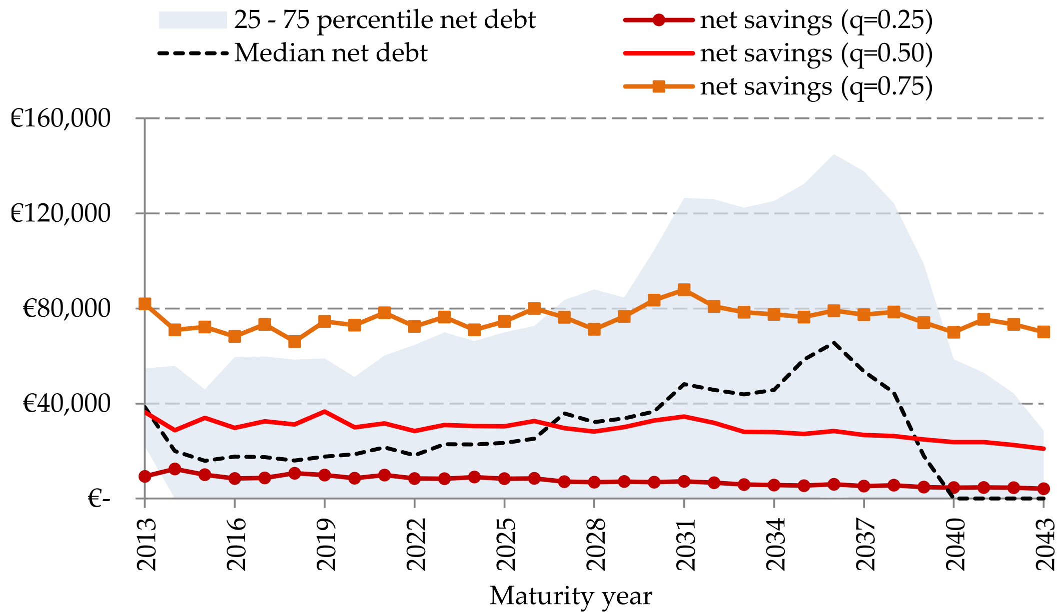

Figure 14: Distribution of net mortgage debt and net savings per maturity year. The quantiles of net savings represent the average of all estimated conditional quantiles of the borrowers with corresponding maturity year. Voluntary repayments are considered, house prices are assumed to remain constant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

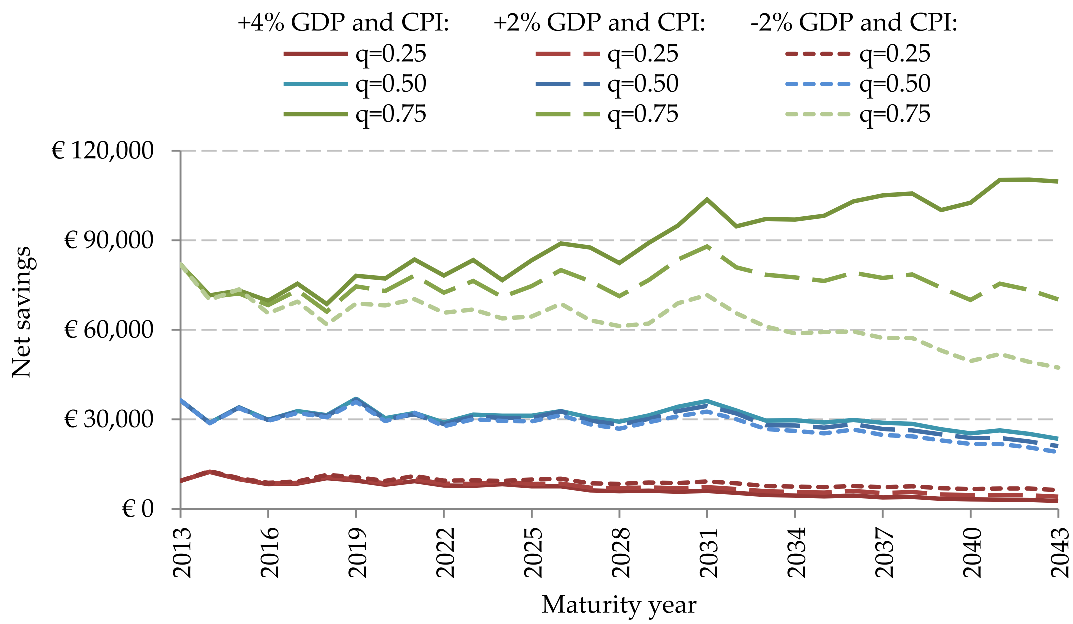

Figure 15: Sensitivity analysis of the distribution of net savings, where different CPI and GDP scenarios are considered in the simulation. The quantiles represent the average of all estimated conditional quantiles of the borrowers with corresponding maturity year.

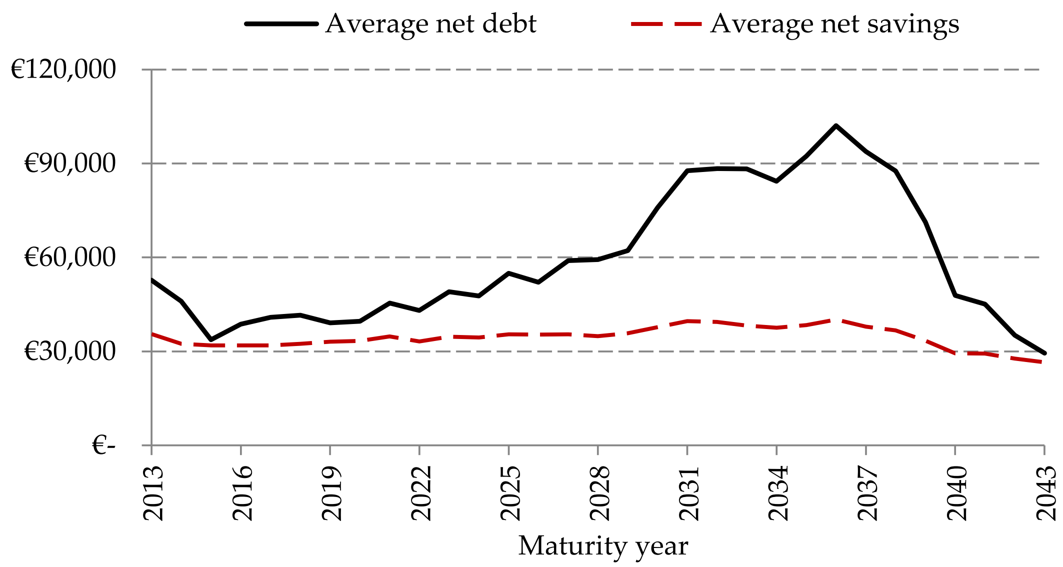

Figure 16: Average net debt and average net savings per maturity year. Here, average net savings are calculated by taking the average of the conditional expectations of all borrowers having a specific maturity year. House prices are considered to remain constant and both GDP and CPI increase with 2%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

6 Discussion

In this thesis we have created a simulation framework that can simulate mortgage debt and non-housing wealth at borrower level. As shown in the previous section, this framework can be used to provide insights on the aggregate mortgage portfolio of the Netherlands. However, its relevance might reach beyond the results presented in this thesis. For example, all lending institutions might be separately investigated in future research. Unfortunately, such results cannot be presented in this thesis due to privacy concerns, but could be helpful to financial supervisors and policymakers.

Moreover, a few remarks and suggestions regarding the quality of the simulation results should be made. First, the possible overestimation of both origination and maturity year might influence the results, not only because they affect the calculations of the contractual repayments. In reality, borrowers might lose their eligibility to tax-deductibility sooner than the reported maturity year. In effect, the simulated borrowers are given more time to make voluntary repayments, meaning we underestimate the mortgage debt at maturity. The correct origination and maturity year should in principle be known to the mortgagee, together with information on previous mortgage loans hold by the borrower. Retrieving this data would substantially improve the quality of the LLD and is highly valuable for future research. Furthermore, once more waves of the LLD will be available, more sophisticated models of voluntary repayments can be considered, exploiting the panel structure of the data. Improving the tractability of borrowers over time (which is not possible for some small banks) and across lending institution will therefore be useful.

Second, the interpretation of the estimated conditional expectation and quantiles is difficult. We might want to consider simulating net savings based on a number of conditional quantiles. That is, we could use the value of the first conditional quantile for a random 1% of the borrowers, the second conditional quantile for a second random 1% of the borrowers, etc. By doing so, we could get simulated data of net savings that is directly comparable to the net debt and therefore much easier to interpret. As a result, we could calculate the exact percentage of borrowers that have not saved enough to repay the mortgage at maturity, together with the actual amount that they fall short. This method does, however, require the estimation of a large number of quantile regressions.

Moreover, in spite of the large number of regressors we find that the models for net savings have a rather poor fit and show only little variation in the fitted values, as can be seen from Figure 14. Future research might consider different matching techniques, such as propensity score matching, if the LLD and the IPO were to be imported in the same environment. Different propensity score methods that are specifically designed to integrating two or more datasets sharing a common subset of covariates are described in D'Orazio et al. (2006) and Rässler (2002).

A final remark regarding the simulation results is that nominal values are used. As we simulate up to thirty years in the future, inflation can have a large effect. We must realize, for instance, that an outstanding debt of 10 000 euro in 2037 is only 6 217 euro expressed in current price level when considering 2% price inflation.

7 Summary & conclusions

This thesis examines the risks associated with the large share of interest-only mortgages in the Dutch mortgage portfolio. Recent reforms have made these mortgages unattractive, but a large legacy from the past has not been subjected to any reform and could possibly stay in the books of banks for ever. 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 to show how risks are distributed across households. We find that 58% of the current net mortgage debt comes from interest-only loans, but that these are often combined with amortizing loans. Borrowers having a full interest-only mortgage are typically older borrowers having substantial home equity, such that the risks regarding these mortgages are limited. We also find that mortgages with high LTV ratios are often backed by the government.

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 much 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 hold by older borrowers. Underwater mortgages are typically amortizing mortgages (at least partly), such that the share of underwater mortgages will decrease rapidly even when house prices remain constant. In more optimistic scenarios, where house prices increase with 3% annually, all mortgages will be above water within ten years. Only if house prices decrease by more than 2% annually and voluntary repayments are not allowed we find an increase in underwater mortgages. Also, almost all mortgages will be above water at maturity when assuming house prices are constant over thirty years. This is a relevant finding for the current discussion on LTV caps.

In this thesis we relate mortgage debt to non-housing wealth (obtained from a different dataset) and show that most households 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 50 000 euro on average). These households could be confronted with an increase in monthly costs as interest payments are then no longer tax-deductible. Nevertheless, almost all borrowers will have a positive home equity at maturity, such that risks associated to the banking sector are limited.

This thesis nuances the gravity of the risks associated with the large portion of interest-only mortgages. Risks are present, but appear to be favorably distributed across households and much of the interest-only loans will still be redeemed over the next thirty year. Most households will not save enough at maturity to fully repay the debt, but risks are only limited as home equity is positive as well. Our conclusions should be taken carefully, as results are sensitive to house price changes and assumptions are needed to make the data usable. The many sensitivity checks though, do not reveal that model-related assumptions drive our results. Data-related assumptions, such as the proxy of inception using origination year, are more crucial. Improving the quality of the data will thus be highly valuable to future research.

The topic of this thesis closely relates to several other issues currently discussed, such as the proposal to further reduce LTV caps below 100% after 2018, the possibility to increase mortgage loans risk weights, or the further selection of high quality mortgages into securitized pools. While we do not specifically aim to contribute to any of these discussions, our results show that LTV reductions take place also thanks to voluntary repayments and that the most common risk triggers appear to positively sorted, which should possibly please the supporters of low risk weights.

Acknowledgment

First of all, I express my deepest thanks to my supervisor Mauro Mastrogiacomo for his excellent guidance and all his work on the loan level data. I am grateful to my supervisor Rob Alessie for all his help and without whom I never have gotten this internship. Also, I am indebted to the entire Financial Stability division at De Nederlandsche Bank for providing me all facilities and freedom that I needed to complete this thesis and, above all, for giving me a wonderful and educational experience.

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.

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.

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). The Risks of the Dutch Mortgage Portfolio, a Micro Econometric Analysis of the DNB Loan Level Data. *DNB Occasional Studies (forthcoming)*.

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

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.

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

Appendices

A Mortgage type definitions

The Dutch mortgage market is characterized by, among other things, its large variety of complex loan structures. In this appendix we outline the most commonly used mortgage products in the Netherlands, although each originator generally has its specific product lines with personal features. In general, all mortgage types can be categorized into three groups: 1) mortgages that start immediately with repaying the mortgage debt, 2) mortgages that defer repayment of the principle until maturity (a so-called "bullet" or "baloon" repayment), and 3) mortgages that never amortize. More specifically, we distinguish the following mortgage types:

1. **Annuity mortgage**
   The periodic payments with an annuity mortgage consist of interest payments and actual repayments of the mortgage. The sum of these two stays the same over the whole period (at least during the fixed interest rate period) and the mortgage will be fully redeemed at maturity. Consequently, the interest payment will decrease over the years, in contrast to the repayment part. As from 2013, new mortgages are only eligible for the interest deduction when the mortgage is fully repaid within thirty years and repayment of the mortgage is done at least according to an annuity scheme. Therefore, most mortgages contracted after 2013 are annuity mortgages.

2. **Linear mortgage**
   This type of mortgage is much similar to the annuity mortgage, only now the mortgage repayment part stays constant over the whole mortgage period. In effect, the amount you pay on interest, and therefore the total monthly costs, will decrease over time. As repayments are done faster compared to the annuity mortgage, linear mortgages are still eligible for interest deduction as well.

3. **Interest-only mortgage**
   With this type of mortgage no repayment scheme of the principal amount is included, but only interest is paid. Moreover, there is no capital accumulated in a mechanism linked to the mortgage, such that the debt service ratio for this type of mortgage is relatively low.

4. **Life insurance mortgage**
   This mortgage consists of a loan and a life insurance policy. During the lifetime of the mortgage there is no repayment of the principle, such that the periodic payments consists of a premium for the insurance part and the interest to be paid on the loan.

The assets accumulated in the life insurance product (in Dutch: Kaptiaalverzekering Eigen Woning (KEW)) must be used to repay the mortgage at maturity or in case of the death of the insured. The capital accumulation in the KEW is tax-free, explaining the popularity of these types of mortgages. In contrast to the savings insurance mortgage, part of the return in the life insurance product depends on the investment returns of the insurer, such that it is uncertain whether you can repay the whole mortgage at maturity. Generally, there is a minimum guaranteed return of about 3%.

5. **Savings insurance mortgage**
   This type of mortgage is similar to the life insurance mortgage, only here you are guaranteed that the accumulated capital (KEW) is enough to repay the mortgage at maturity.

6. **Investment mortgage**
   This type is again similar to the life insurance and savings insurance mortgage, only now the returns on the capital in the life insurance product are completely based on the returns of an investment portfolio. Usually, the mortgagee can choose from a wide range of investment funds with different risk profiles. Since the returns of the investment portfolio are not known in advance, you are again not guaranteed that you can pay off the whole loan at maturity.

7. **Bank savings mortgage**
   Instead of in a life insurance product, this type of mortgage accumulates capital in either a blocked savings (SEW) or investment (BEW) account at the bank, both linked to the mortgage. The SEW and BEW are the banking counterparts of the KEW, where the main difference is that the capital is accumulated at the same institution as where the mortgage loan is accommodated and no life insurance is included (although this can be arranged separately). As from 2008 the SEW and BEW are allowed for favorable tax treatment as well, while before 2008 only insurance products offered this tax benefit. This change in tax system resulted in an increase in popularity of the bank savings mortgage, competing with the savings insurance and investment mortgage.

8. **Hybrid mortgage**
   This mortgage is a combination between a savings insurance mortgage and an investment mortgage. During the lifetime of the mortgage one can switch between a guaranteed return on the life insurance product, or a return including investment risk.

The RMBS template from the ECB, which is used to collect the loan-level data, is naturally not specifically designed for the Dutch mortgage market, but for all European countries in general. Consequently, the dataset does not always allow for a perfect identification

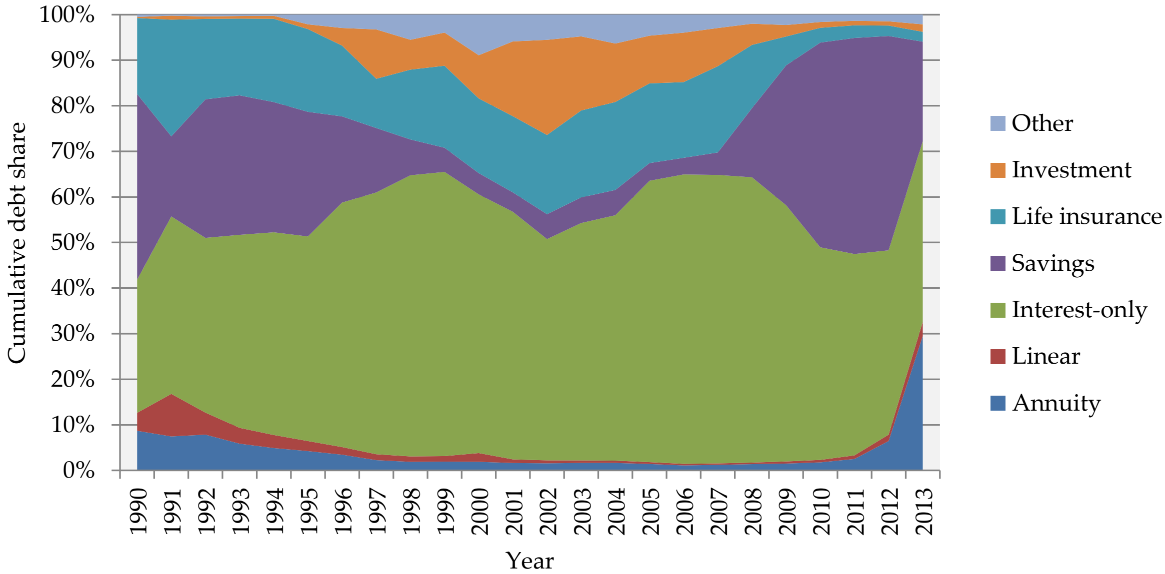

Figure 17: Debt-weighted share of all new and renegotiated mortgages per mortgage type and origination year, as defined in the LLD.

of the mortgage type. Specifically, in the loan-level data a mortgage can be classified as:

1) Annuity,
2) Linear,
3) Bullet,
4) Bullet + Savings deposit,
5) Bullet + Life insurance,
6) Bullet + Investments, or
7) Other.

Fortunately, the most relevant mortgage type in our analysis (interest-only) is perfectly identified as a bullet mortgage. Nor is there any doubt in identifying annuity and linear mortgages. However, for some other mortgage types in the Netherlands more than one of the above options could apply. For example, an investment mortgage has a life insurance, and the capital accumulated in that life insurance is based on investment returns. In the same line of reasoning we can classify a savings insurance mortgage as either option 4 or 5 in the list above. This could impose some complications, especially when different definitions are used between (or within) institutions. By looking at Figure 17 we see that the scale of this problem is limited, although some remarks should be made. The figure displays the popularity of the different mortgage types as defined in the LLD over time. For example, we observe a steep increase in the popularity of annuity mortgages in 2013, in accordance with current policy. Interestingly, around 2008 we observe a steep increase in popularity

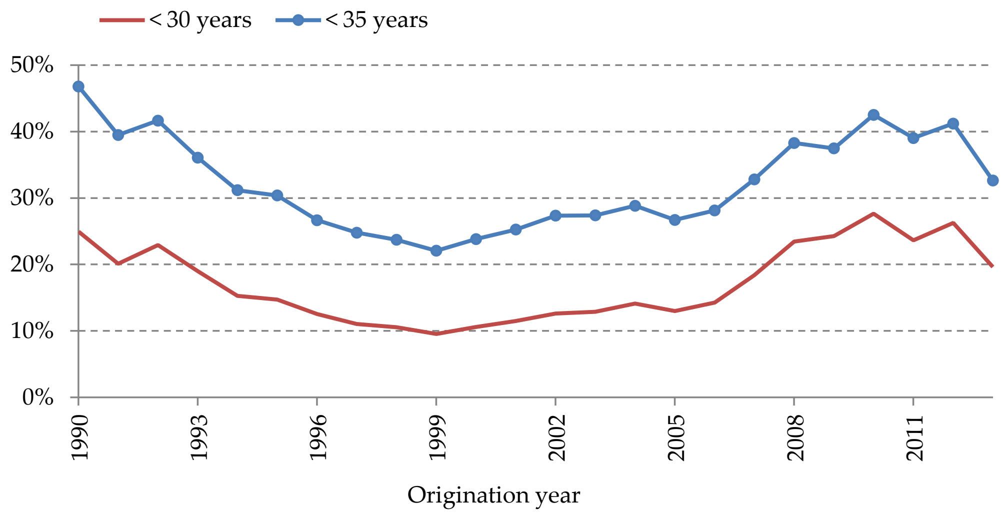

Figure 18: Share of borrowers that are younger than a specific age at loan origination, as defined in the LLD.

for the mortgage type classified as *Bullet + Savings deposit*, indicating that this category includes the bank savings mortgages. However, bank savings mortgages were definitely not that popular in the 1990s as indicated by the figure, meaning that more mortgage types fall under this category. We presume that savings insurance mortgages are sometimes classified as *Bullet + Savings deposit* and sometimes as *Bullet + Life insurance*. Unfortunately, no consistent patterns are found in the classifications such that no perfect identification is possible. In our microsimulation model, however, we treat savings insurance mortgages, life insurance mortgages and bank savings mortgages all the same, such that this misclassification does not pose a large problem. Finally, the hype of investment mortgages (a.k.a. *woekerpolis*) around 2000 is clearly observed in the figure as well, after which the popularity of this type of mortgage decreased due to very disappointing investment returns.

B Data complications

Generally, we would expect that almost all starters (first-time home buyers) are younger than thirty-five years. However, as can be seen in Figure 18, on average roughly 20% and 30% of all borrowers in the loan-level data is younger than respectively thirty and thirty-five years at first loan origination. This is because the information at loan origination is updated whenever a loan is re-contracted, such that accurate information on loan characteristics at origination is lost. Realizing, for instance, that most households move houses more than once (associated with mortgage renegotiations) the small percentage of young starters makes more sense. From Figure 18 we observe that the share of young borrowers is particularly low for the period around 2000, which could be seen as an indication that more

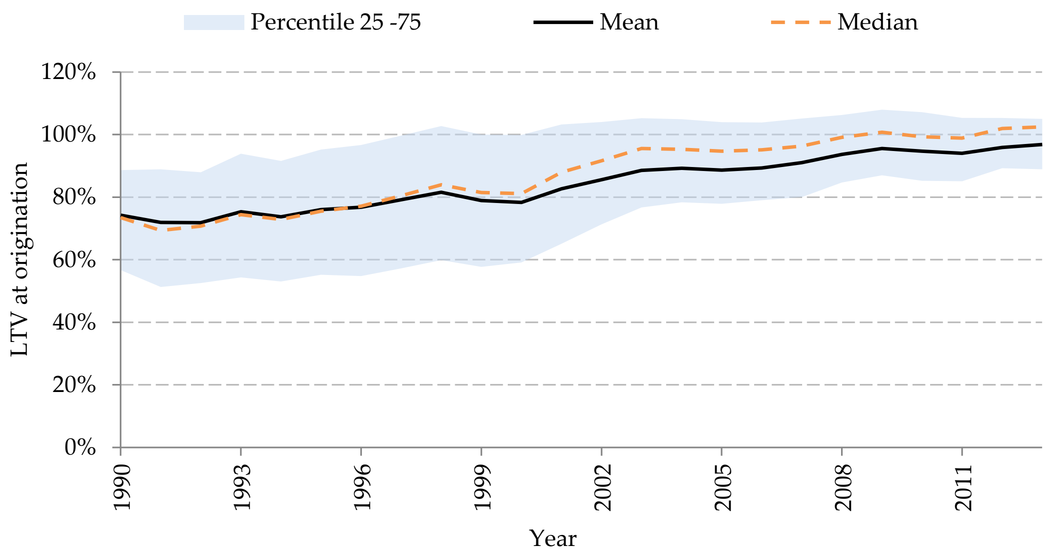

Figure 19: Distribution of the LTV for first-time home buyers per origination year.

mortgage renegotiations have taken place. House prices increased rapidly prior and during this period. Mortgage renegotiations were popular to cash the increased home equity, for instance by taking an interest-only mortgage.

These issues call for a careful interpretation of the data. As an example, Figure 19 presents the evolution of the LTV for starters, based on the most recent wave of the LLD. Here, we define starters as being younger than thirty years at first loan origination, as other borrowers are likely to have renegotiated their loans and therefore do not provide accurate information on their first mortgage. Along the same line of reasoning we only include borrowers for which all loans are originated in the same year, as in other cases we simply do not know whether the more recent loans are renegotiated or not. The resulting figure seems reliable; we observe an increase in average original LTV in the 1990s until the beginning of the last decade, where an average original LTV of 95% has become the norm in more recent years. Notice that in 2012 and 2013 the 75 percentile lies below the LTV cap of 106% and 105%, respectively. Again, we have to realize that the data might still include some re-contractors instead of starters, thereby underestimating the LTV for starters. On the other hand, actual starters older than thirty years at origination probably have had a lower LTV as well, but are excluded in the figure. Moreover, the original LTV might be overestimated since we observe the original LTV conditionally on being observed in 2013. Borrowers that have paid off their whole mortgage probably have had a lower LTV at origination. This effect is limited, however, as a mortgage is typically contracted for thirty years. A final note regarding the calculations of the original LTV is that the value at origination is approximated using the Dutch house price index calculated by Statistics Netherlands, as we only observe the most recent valuation amount of the property.

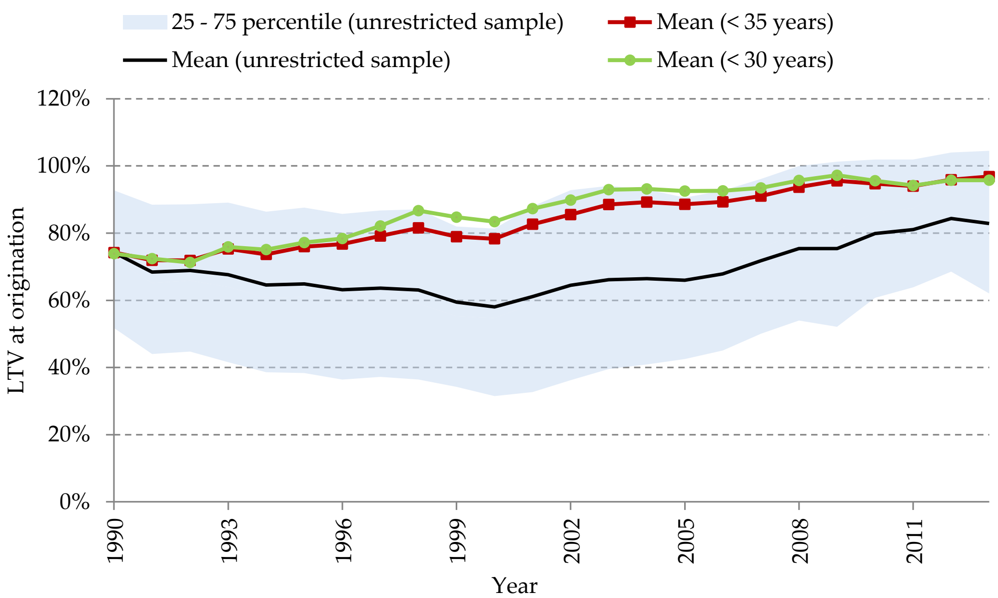

Figure 20: Average LTV at origination using different age limits to define a first-time home buyer.

Figure 20 presents the average LTV at origination over time for different definitions of a starter. As can be seen, without imposing an age limit the results differ dramatically due to the fact that we have included re-contractors, who typically have a lower LTV. Especially for the period around 2000, for which we argued that most renegotiations have taken place, the original LTV is unrealistically low. Using only borrowers that are younger than thirty or thirty-five years at origination yields a much more realistic figure. The difference between these two additional figures is minimal, only for the period around 2000 there is a small difference visible. This robustness check reassures that imposing an age limit of thirty years should yield a rather good approximation of the historic development of the LTV at origination.

C Asymptotically equivalent Lagrange Multiplier test for heteroskedasticity

Consider the latent model

$$y_i^* = \boldsymbol{x}_i'\gamma + \varepsilon_i, \ i = 1, 2, \dots, N,$$

where y_i^* is the latent response variable underlying y_i, $\boldsymbol{x}_i$ is an $(k+1) \times 1$ vector of a constant and k explanatory variables for which $\gamma = (\gamma_1, \dots, \gamma_{k+1})'$ are the corresponding coefficients. Moreover, $\varepsilon_i \sim NID(0, \sigma^2)$ and independent of $\boldsymbol{x}_i$. Now, instead of observing the latent variable y_i^*, we observe

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

such that w_i follows a probit model. When estimating a probit model for w_i we have to assume that $V(\varepsilon_i) = \sigma^2 = 1$, as otherwise the model is not identified. Denote the estimate of γ resulting from a probit regression by $\widehat{\gamma}_r$. Alternatively, consider $V(\varepsilon_i) = g(\boldsymbol{z}_i'\boldsymbol{\alpha})$, for some function $g(.) > 0$ with $g(0) = 1$, $g'(0) \neq 0$ and where the $k \times 1$ vector $\boldsymbol{z}_i$ contains all elements of $\boldsymbol{x}_i$ except the constant. Now, a test for heteroskedasticity boils down to testing $\boldsymbol{\alpha} = \boldsymbol{0}$.

First, let us define the $(2k+1) \times 1$ vector $\boldsymbol{\delta} = (\gamma', \boldsymbol{\alpha}')'$ and denote its estimate under $H_0 : \boldsymbol{\alpha} = \boldsymbol{0}$ by $\widehat{\boldsymbol{\delta}}_r = (\widehat{\gamma}_r', \boldsymbol{0}')'$. The probability mass function of w_i is given by

$$f(w_i|\boldsymbol{x}_i) = \Phi\left(\frac{\boldsymbol{x}_i'\gamma}{g(\boldsymbol{z}_i'\boldsymbol{\alpha})}\right)^{w_i} + \left(1 - \Phi\left(\frac{\boldsymbol{x}_i'\gamma}{g(\boldsymbol{z}_i'\boldsymbol{\alpha})}\right)\right)^{1-w_i},$$

where $\Phi()$ denotes the standard normal cumulative distribution function. Given independence over i we get the following log-likelihood function:

$$\ln L(\boldsymbol{\delta}) = \sum_{i=1}^{N} \left\{ w_i \ln \Phi\left(\frac{\boldsymbol{x}_i'\gamma}{g(\boldsymbol{z}_i'\boldsymbol{\alpha})}\right) + (1 - w_i) \ln\left(1 - \Phi\left(\frac{\boldsymbol{x}_i'\gamma}{g(\boldsymbol{z}_i'\boldsymbol{\alpha})}\right)\right) \right\}.$$

Now, the derivative of $\ln L(\boldsymbol{\delta})$ w.r.t. $\boldsymbol{\delta}$ determined at $\widehat{\boldsymbol{\delta}}_r$ is given by

$$\left.\frac{\partial \ln L(\boldsymbol{\delta})}{\partial \boldsymbol{\delta}}\right|_{\boldsymbol{\delta}=\widehat{\boldsymbol{\delta}}_r} = \begin{pmatrix} \left.\frac{\partial \ln L(\gamma, \boldsymbol{\alpha})}{\partial \gamma}\right|_{\boldsymbol{\delta}=\widehat{\boldsymbol{\delta}}_r} \\ \left.\frac{\partial \ln L(\gamma, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right|_{\boldsymbol{\delta}=\widehat{\boldsymbol{\delta}}_r} \end{pmatrix},$$

where

$$\left.\frac{\partial \ln L(\gamma, \boldsymbol{\alpha})}{\partial \gamma}\right|_{\boldsymbol{\delta}=\widehat{\boldsymbol{\delta}}_r} = \sum_{i=1}^{N} \left\{ \frac{w_i}{\Phi\left(\boldsymbol{x}_i'\widehat{\gamma}_r\right)} \phi\left(\boldsymbol{x}_i'\widehat{\gamma}_r\right) \boldsymbol{x}_i - \frac{1 - w_i}{1 - \Phi\left(\boldsymbol{x}_i'\widehat{\gamma}_r\right)} \phi\left(\boldsymbol{x}_i'\widehat{\gamma}_r\right) \boldsymbol{x}_i \right\} = \sum_{i=1}^{N} \widehat{\epsilon}_i \boldsymbol{x}_i,$$

and

$$\left.\frac{\partial \ln L(\gamma, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right|_{\boldsymbol{\delta}=\widehat{\boldsymbol{\delta}}_r} = \sum_{i=1}^{N} \left\{ \frac{w_i}{\Phi\left(\boldsymbol{x}_i'\widehat{\gamma}_r\right)} \phi\left(\boldsymbol{x}_i'\widehat{\gamma}_r\right) \kappa \boldsymbol{x}_i'\widehat{\gamma}_r \boldsymbol{z}_i - \frac{1 - w_i}{1 - \Phi\left(\boldsymbol{x}_i'\widehat{\gamma}_r\right)} \phi\left(\boldsymbol{x}_i'\widehat{\gamma}_r\right) \kappa \boldsymbol{x}_i'\widehat{\gamma}_r \boldsymbol{z}_i \right\} = \sum_{i=1}^{N} \kappa \widehat{\epsilon}_i \boldsymbol{x}_i'\widehat{\gamma}_r \boldsymbol{z}_i,$$

where the generalized residuals $\widehat{\epsilon}_i$ are implicitly defined and $\kappa = -g'(0)$ is an irrelevant constant.

Now, define the $(2k+1)\times 1$ vector $\widehat{\boldsymbol{s}}_i = (\widehat{\epsilon}_i \boldsymbol{x}_i', \widehat{\epsilon}_i \boldsymbol{x}_i' \widehat{\gamma}_r \boldsymbol{z}_i')'$. Following (Cameron and Trivedi, 2005) we have that an asymptotically equivalent version of the Lagrange Multiplier (ML) test statistic is given by

$$LM^* = \left(\sum_{i=1}^{N} \widehat{\boldsymbol{s}}_i'\right)\left[\sum_{i=1}^{N} \widehat{\boldsymbol{s}}_i \widehat{\boldsymbol{s}}_i'\right]^{-1}\left(\sum_{i=1}^{N} \widehat{\boldsymbol{s}}_i\right),$$

which has an χ^2 distribution with k degrees of freedom under the null hypothesis of homoskedasticity. Finally, this statistic can be computed using an auxiliary regression of $\boldsymbol{\iota}$ on $\widehat{\boldsymbol{s}}_i$, where $\boldsymbol{\iota}$ is a $N\times 1$ vector of ones. Now, let $\boldsymbol{S}$ be the $N \times (2k+1)$ matrix with rows $\widehat{\boldsymbol{s}}_i'$, such that we can write

$$\begin{aligned} LM^* &= \boldsymbol{\iota}'\boldsymbol{S}\left(\boldsymbol{S}'\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\boldsymbol{\iota} \\ &= N\frac{\boldsymbol{\iota}'\boldsymbol{S}\left(\boldsymbol{S}'\boldsymbol{S}\right)^{-1}\boldsymbol{S}'\boldsymbol{\iota}}{\boldsymbol{\iota}'\boldsymbol{\iota}} \\ &= NR_u^2, \end{aligned}$$

where R_u^2 is the uncentered R^2 from the auxiliary regression of $\boldsymbol{\iota}$ on $\widehat{\boldsymbol{s}}_i$.

D Extended estimation results

	0% I-O		20% I-O		40% I-O	
Regressors	Coef	ME	Coef	ME	Coef	ME
age/10	0.229***	-0.000674	-0.0625	0.000561	0.282***	0.0134***
	(0.0165)	(0.000519)	(0.0411)	(0.00156)	(0.0264)	(0.000809)
$(\text{age}/10)^2$	-0.0269***		0.00454		-0.0271***	
	(0.00164)		(0.00420)		(0.00274)	
share I-O	0.778***	0.140***	0.410***	0.096***	-0.624***	-0.124***
	(0.147)	(0.0264)	(0.0897)	(0.0211)	(0.0616)	(0.0122)
interest rate	-1.114***	-0.200***	-0.282	-0.0664	-2.489***	-0.495***
	(0.258)	(0.0464)	(0.684)	(0.161)	(0.469)	(0.0933)
underwater	-0.532***	-0.0524***	-0.584***	-0.0115**	-0.488***	-0.0353***
	(0.0298)	(0.00187)	(0.0610)	(0.00451)	(0.0361)	(0.00226)
age × underwater	0.00530***		0.0119***		0.00716***	
	(0.000764)		(0.00148)		(0.000909)	
NHG	-0.204***	-0.0367***	-0.0835***	-0.0196***	-0.207***	-0.0411***
	(0.00597)	(0.00107)	(0.0118)	(0.00278)	(0.00756)	(0.00150)
current $\text{LTV}/10^2$	0.220***	0.0395***	-0.294***	-0.0692***	-0.143***	-0.0284***
	(0.0109)	(0.00197)	(0.0248)	(0.00583)	(0.0172)	(0.00343)
Constant	-1.652***		-0.636***		-1.160***	
	(0.0441)		(0.111)		(0.0746)	
N	434 327		85 912		236 161	
Log-likelihood	-144 160		-36 567		-85 923	
pseudo R^2	0.014		0.010		0.032	

	60% I-O		80% I-O		100% I-O	
Regressors	Coef	ME	Coef	ME	Coef	ME
age/10	0.324***	0.00586***	0.200***	-0.00488***	0.159***	-0.0276***
	(0.0248)	(0.000789)	(0.0303)	(0.000901)	(0.0113)	(0.000355)
$(\text{age}/10)^2$	-0.0318***		-0.0230***		-0.0244***	
	(0.00248)		(0.00281)		(0.00963)	
share I-O	0.097*	0.022*	0.498***	0.115***	-1.351***	-0.317***
	(0.0554)	(0.0123)	(0.0692)	(0.0160)	(0.121)	(0.0283)
interest rate	-3.041***	-0.675***	-1.352***	-0.313***	4.812***	1.129***
	(0.411)	(0.0913)	(0.462)	(0.107)	(0.216)	(0.0507)
underwater	-0.0715*	-0.0159***	-0.349***	-0.0163***	-0.712***	0.0993***
	(0.0383)	(0.00243)	(0.0460)	(0.00317)	(0.0417)	(0.00299)
age × underwater	-.00000644e		0.00547***		0.0196***	
	(0.000887)		(0.000976)		(0.000806)	
NHG	-0.0373***	-0.00828***	0.0881***	0.0204***	0.148***	0.0348***
	(0.00788)	(0.00175)	(0.0131)	(0.00304)	(0.00924)	(0.00217)
current $\text{LTV}/10^2$	-0.213***	-0.0472***	-0.260***	-0.0603***	-0.285***	-0.0668***
	(0.0151)	(0.00336)	(0.0179)	(0.00414)	(0.00771)	(0.00181)
Constant	-1.585***		-1.548***		0.175	
	(0.0723)		(0.100)		(0.127)	
N	243 213		165 142		736 811	
Log-likelihood	-98 228		-69 312		-312 879	
pseudo R^2	0.006		0.009		0.013	

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 8: Probit regression (Part I of log-normal hurdle) on voluntary repayments (1 = voluntary repayment, 0 = no voluntary repayment) per interest-only category.

	0% I-O	20% I-O	40% I-O	60% I-O	80% I-O	100% I-O
Regressors	Coef	Coef	Coef	Coef	Coef	Coef
age	0.0553***	0.0289***	0.0363***	0.0603***	0.0535***	0.0283***
	(0.00323)	(0.00766)	(0.00517)	(0.00511)	(0.00583)	(0.00183)
$(\text{age}/10)^2$	-0.0547***	-0.0235***	-0.0297***	-0.0560***	-0.0543***	-0.0279***
	(0.00323)	(0.00782)	(0.00531)	(0.00511)	(0.00546)	(0.00159)
share I-O	-4.305***	0.562***	0.0569	0.358***	0.193	2.534***
	(0.263)	(0.165)	(0.116)	(0.107)	(0.128)	(0.191)
interest rate	-3.727***	-7.001***	-7.243***	-7.936***	-9.365***	-10.79***
	(0.530)	(1.223)	(0.900)	(0.835)	(0.914)	(0.379)
underwater	-0.866***	-0.929***	-0.693***	-0.115	-0.712***	-1.051***
	(0.0582)	(0.115)	(0.0728)	(0.0772)	(0.0881)	(0.0647)
age $\times$ underwater	0.0123***	0.0188***	0.0125***	0.000225	0.00929***	0.0173***
	(0.00148)	(0.00278)	(0.00181)	(0.00178)	(0.00187)	(0.00123)
NHG	-0.278***	-0.104***	-0.0833***	0.163***	0.155***	0.00973
	(0.0113)	(0.0217)	(0.0145)	(0.0152)	(0.0232)	(0.0140)
current LTV	0.00891***	0.00733***	0.00736***	0.00731***	0.00793***	0.0112***
	(0.000206)	(0.000425)	(0.000308)	(0.000276)	(0.000312)	(0.000125)
Constant	7.910***	8.425***	8.387***	7.736***	8.188***	6.364***
	(0.0850)	(0.205)	(0.145)	(0.146)	(0.190)	(0.202)
N	45 758	13 251	29 384	34 253	24 844	113 744
R^2	0.073	0.045	0.036	0.039	0.048	0.116
Log-likelihood	-65 782	-19 975	-44 903	-52 985	-38 034	-158 957
$\widehat{\sigma}^2$	1.019	1.093	1.116	1.137	1.119	0.979
Two-Part model:						
pseudo R^2	0.013	0.004	0.016	0.002	0.004	0.010
Log-likelihood	-642 270	-182 005	-411 212	-482 717	-347 965	-1 560 000

Standard errors in parentheses
*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 9: Part II of log-normal hurdle per interest-only category (OLS regression on the natural logarithm of voluntary repayments for only those observations with a positive voluntary repayment).