Determinants of Household Borrowing in the United States and a Panel of OECD Countries Prior to the Financial Crisis of 2007/2008

Rafael WILDAUER

September 2017

Department of Economics

Kingston University

Table of Content

Acknowledgements

Finishing this piece of work is a suitable moment to say thank you for all the support, encouragement, friendship and love I have received from several people. First of all I thank my supervisors Engelbert Stockhammer and Chris Stewart for taking the time to discuss my ideas and findings as well as providing excellent advice when it was needed. Based on the PhD experience of some of my friends and relatives, I am aware that the interest in my research as well as the respectful and open-minded discussions I had especially with my first supervisor, cannot be taken for granted. I did not plan to do a PhD until the end of my Master degree and on that occasion I want to thank two people in particular who encouraged me to pursue that route. My uncle Helmut Ettl introduced me not only to Kaleckian macroeconomics but also strongly supported my decision to seek a PhD position. My friend Laura Dobusch was a very important early voice which got me to start thinking about a PhD. Special thanks also goes to Jakob Kapeller who gave me an early opportunity to gather research experience at the University of Linz which made handling the Survey of Consumer Finances much easier and also greatly helped in successfully entering the job market. The decision of pursuing a PhD was not only highly rewarding from an academic perspective but also resulted in meeting outstanding people: Most importantly my PhD colleague Javier López Bernardo, with whom I shared many delicious pizzas and had a lot of fun and who is a very knowledgeable conversation partner to discuss my research and to collaborate with beyond the PhD. While professional support is important it is at least as important to be able to rely on the support of family and friends. Here my special thanks goes to my old friend Stefan Schreiner who is simply the best friend one can ask for. For the last 19 years he was always there to cheer me up, joke around and give advice about life in general. A special thank you goes to my parents, Margit and Rudolf Wildauer. After settling the conflict about whether or not I should attend my high school graduation ceremony, their love and support was always there. Last but not least I want to thank Arianne Griffith for enriching my life and making it so much more worth living on top of helping and supporting me in achieving my goals. I am incredibly lucky to be with her. Finally I want to thank my family, my friends and my colleagues whom I could not mention here individually. Without all of you the last 42 months would not have been as interesting, fun and exciting.

Abstract

This thesis investigates which factors drove the spectacular accumulation of household liabilities prior to the financial crisis of 2007/2008 in the United States and 12 other OECD countries. Two mechanisms are of particular interest. The first is the polarization in the distribution of income which can be observed since the 1980s in the United States and European OECD countries (Atkinson et al. 2011). The second mechanism is the increase in asset prices and in particular residential real estate prices, which is considered by some authors as a major explanation of household debt accumulation (Mian & Sufi 2011). Based on two different data sources, the Survey of Consumer Finances (SCF) and a macro panel of 13 OECD countries, an extremely robust result emerges: The residential housing market is the key driver of household debt accumulation. This finding strongly supports an asset-focused view of household sector debt and discourages explanations which rely on a direct link from the distribution of income to household indebtedness. There is some evidence that the income polarization contributed to higher debt-levels in the US but this positive relationship is conditional on homeownership. The interpretation of these findings is that real estate purchases represent the single most important reason for households to take on debt and even if households take on debt for other reasons, real estate collateral is often a binding requirement to be granted a loan. The thesis contributes to the existing literature in three ways. First, it is the first attempt to investigate the inequality as well as the asset mechanism in a unified framework whereas previous studies analysed them in isolation. Second, by constructing a borrowing measure from SCF data the unmatched coverage of the top tail of the wealth distribution of this dataset can be exploited. Third, the thesis provides some methodological insights when using survey data to analyse household borrowing behaviour: increased model fit by the separation of borrowing and non-borrowing households; necessity to separately control for asset purchases; disadvantages of growth rates and logarithmic differences.

1 Introduction

Debt bubbles have the potential to cause severe recessions. Since the financial crisis of 2007/2008 there is a growing body of literature which documents that private sector indebtedness in general and household debt in particular is a good predictor of financial crises (Schularick & Taylor 2012; Bezemer & Zhang 2014; Borio 2014; Eichengreen & Mitchener 2003) and there is some evidence that expanding credit to the household sector and especially to low quality borrowers will eventually trigger mass defaults and adverse growth effects when asset prices start to decline[1] (Bezemer et al. 2014; Mian & Sufi 2009). Many other examples for the potentially harmful effects of private sector debt accumulation like the Asian crisis at the end of the 1990s, the Great Depression in the 1930s and the many banking crisis in the (long) 19th century (1830s, 1857, 1873, 1893, 1907) can be found. In addition to the historic evidence the destructive potential of debt bubbles is also recognized on theoretical grounds (Bernanke et al. 1998; Fisher 1933; Koo 2011; Minsky 1978). Bernanke et al. (1998) for example propose a financial accelerator mechanism which is based on the assumption that bursting debt bubbles, depress asset prices and thus borrowers' balance sheets. However, with less skin in the game the potential conflict of interest between the bank and the borrower increases and thus market interest rates go up, depressing output and production. Fisher (1933) argues that excessive debt loads lead to bankruptcies and in a deflationary environment this increases the real burden of debt which leads to further bankruptcies which depresses output and increases the deflationary pressure. Koo (2011) argues that over indebted firms start to prioritize debt minimization instead of profit maximization and thus cut investment spending. The resulting drop in expenditures leads to a severe output contraction. Minsky's Financial Instability Hypothesis (1978) received renewed attention after the financial crisis of 2007. He argues that stability breads instability as the riskiness of financing structures increases in periods of economic prosperity. The eventual collapse of the bubble triggers a severe recession.

There is an alternative argument which states that the primary objective for preventing deep recessions is to ensure a panic proof design of the financial system (Ricks 2016). If this objective is achieved debt bubbles are less worrying because without a meltdown of the financial sector they will not develop into deep recessions. While it is important to think about how to design the financial institutional framework (which is another term for financial regulation) in order to limit the damage the financial sector inflicts on the broader economy in case of a crisis, bursting bubbles are still likely to damage the economy even in a panic proof system. This means improving the institutional design of the financial sector and employing policies which are aimed at preventing or limiting bubbles from building are complementary and not conflicting measures. This thesis focuses

[1] The less sharp decline in property prices and less aggressive lending to low income households compared to the US are the two main reasons why the UK so far managed to cope with household debt levels well above 100% of disposable income.

on the second part and leaves questions about financial regulation for future research. Readers interested in current developments of financial regulation in the US and the shortcomings of the recent reforms will find Ricks (2016) and The Volcker Alliance (2016) to be useful starting points.

Against the background of the empirical evidence and the cited theoretical arguments it is of major interest to understand which factors lead to the accumulation of private debt to design policies which are aimed at preventing excessive debt accumulation. The goal of such policies is to limit the damage debt bubbles cause to society in lost output, increased unemployment and human suffering in general. While all theories mentioned in the first paragraph are concerned with the liabilities of the corporate sector, the recent crisis of 2007/2008, which triggered the deepest recession in the US after the Second World War, was closely connected to the accumulation of liabilities in the household sector. The events which began to unravel in 2007 forcefully demonstrated that household sector debt can be the driving factor of macroeconomic outcomes. It is this shocking but recent experience which motivates the analysis of household rather than corporate debt in my thesis.

While there are theoretical (Bernanke et al. 1998; Fisher 1933; Koo 2011; Minsky 1978) as well as empirical arguments (Schularick & Taylor 2012; Bezemer & Zhang 2014; Borio 2014; Eichengreen & Mitchener 2003; Bezemer et al. 2014; Mian & Sufi 2009) which link household debt to financial stability, the factors which drive such an accumulation are less clear. Most specifically, why do households decide to take on debt levels which they eventually fail to handle? The textbook version of the life-cycle consumer would use debt only to smooth consumption over her lifespan and never default on it. Thus, explanations going beyond the standard consumption model are needed. Using different data sources and different methods this thesis investigates three popular explanations for why (US) households became heavily indebted prior to the financial crisis.

The first explanation under assessment is called the expenditure cascades hypothesis (ECH) and is the reformulation by Frank et al. (2014) of an old idea which can be traced back at least to Veblen (1899) and Duesenberry (1949): households care about their social status and their position in society relative to others. In particular if people compare themselves to richer peers then rapidly growing top incomes will trigger a debt financed cascade of expenditures in an effort by lower income households to preserve their relative position. The phenomenon is more widely known as ‘keeping up with the Joneses’. Thus the expenditure cascade hypothesis considers rising income inequality as the main driver of household indebtedness. Several prominent authors have recently stressed the link between rising income inequality and household indebtedness (Rajan 2010; Van Treeck 2014; Kumhof & Ranciere 2010).

However the arguments involved in linking income inequality and indebtedness are quite diverse. Barba and Pivetti (2009) argue that rising inequality has pushed poor households into debt as they were trying to maintain their consumption level. Kumhof and Ranciere (2010) make a similar argument in a DSGE framework. This second explanation will be referred to as the income stagnation hypothesis (ISH). According to the ISH the main reason for the rise in household indebtedness is that in the face of stagnant or even falling real incomes, households rely on debt to prevent living standards from falling. Similar to the ECH, the ISH identifies increasing income inequality as the main factor for rising household debt. The key difference between these two hypotheses is that the ISH relies on path dependent, self-regarding consumption norms, whereas the ECH relies on upward looking other-regarding consumption norms. While their predictions regarding household debt are similar and they ultimately agree that the increase in debt is based on the growth of expenditures exceeding the growth of income, they differ in their predictions regarding consumption expenditures. While the ECH is in line with rising consumption to GDP ratios the ISH predicts stable or stagnant consumption to GDP ratios as households are only able to partially substitute income by taking on debt[2].

Table 1: Investigated hypothesis

label	main cause of rising household debt
expenditure cascades hypothesis (ECH)	increasing top incomes trigger debt-financed status-driven expenditures which cascade down the income distribution
income stagnation hypothesis (ISH)	stagnant or falling real incomes require borrowing to maintain living standards
housing hypothesis (HH)	rising house prices requiring first time buyers taking out larger mortgages and making debt financed purchases attractive in the light of expected price increases

The third explanation will be labelled the housing hypothesis because it rests on the simple idea that house price increases require first time buyers to take on larger mortgages if they are not willing to postpone their purchase. In addition, homeowners will take on additional debt in order to cash in on capital gains and speculators will buy debt-financed property to enjoy future capital gains. The housing hypothesis is supported by a straightforward observation: most household debt is, in fact, mortgage debt. Borio (2014), Goodhart and Hofmann (2008) and Leamer (2007) identify property prices as one of the key variables for financial and business cycles. Ryoo (2015) formalises these ideas and presents a Minsky model where household debt is driven by property prices. However while there is an extensive literature on the effects of property prices or housing wealth on consumption expenditures (see the surveys by Cooper and Dynan (2016) and Paiella (2009) and

[2] This prediction is based on the assumption of reasonably stable average propensities to consume. In general one can think of a situation in which aggregate consumption rises despite falling real incomes and only modestly increasing household liabilities, if households strongly increase their consumption propensities (i.e. lower their saving rate).

references therein), there is much less work on their effects on household debt. The three hypotheses and the corresponding explanations of rising household debt are summarised in Table 1.

An empirical assessment of the explanatory power of these three hypotheses also needs to consider other factors which affect household borrowing decisions in order to avoid omitted variable problems. The first such factor is the role of the financial sector and its willingness to lend. Loans in general and mortgages in particular need to be approved by the borrower's bank and thus credit supply conditions can be a binding constraint independent of the debtor's motivations to take on debt. If the financial sector becomes more willing to accept customers with low credit ratings and accepts higher loan to value ratios, household sector indebtedness will rise. The second factor, for which the empirical analysis controls for, are interest rates. If central banks keep interest rates at very low levels, cheap (mortgage) rates will stimulate household borrowing. Taylor (2009) prominently argues that low interest rates contributed to the financial crisis and led to an over indebted household sector. Brancaccio and Fontana (2013) argue that a negative relationship between real interest rates and economic activity and borrowing can not only be derived from a Taylor rule but also from a 'solvency rule' which makes the impact monetary policy has on indebted households and firms explicit. Either way the interest rate has an important impact on household finances not only by attracting or deterring new borrowing but also by determining the flow of interest payments stemming from already existing liabilities. Especially in a situation with an already large amount of outstanding debt, small changes in interest rates can have large effects on the related interest payments. This latter effect depends to a large extent on the amount of outstanding liabilities, which is the third factor the empirical analysis controls for. The reasoning is that a large amount of outstanding debt makes households financially more vulnerable and thus discourages additional borrowing. However using a large fraction of household disposable income for interest and principal (re)payments, might encourage borrowing for other purposes like consumption or equity purchases because less income is available to finance these expenditures. This leads to the fourth factor: disposable household income. At the individual level and with a given set of expenditures, having more income available requires less borrowing to fund these expenditures. However as income rises also the standard of living increases and higher incomes allow to serve larger loans (mortgages). Therefore higher incomes are most likely related to higher levels of household borrowing. In addition, demographic factors are considered, most importantly the age structure of the population. If households attempt to smooth consumption expenditures over the life-cycle, changes in the age structure might have important implications for household saving and borrowing at the aggregate level. An increasing share of old-age consumers, in particular, should lead to less borrowing as these consumers will run down their assets in the absence of bequest motives.

So far the picture I have painted is a gloomy and negative one: Household sector debt is a likely source of financial instability and recessions. This is only one part of the picture. At the same time accumulation of household debt can make significant contributions to economic growth if it is used to finance consumption or residential construction expenditures. There is empirical evidence that accumulation of household debt was indeed a major driver of economic growth in many OECD countries (Stockhammer & Wildauer 2016). It is against this background of household debt as a source of financial and economic instability on the one hand and as a stimulant of economic growth on the other hand that I analyse the determinants of household borrowing: If debt accumulation is primarily the result of inflated asset prices (most importantly residential real estate) due to inelastic supply, the aggregate demand effect is likely to be small in relation to the amount of debt taken on. In contrast if debt is primarily used to finance consumption expenditures either due to social status considerations or in an attempt to maintain a certain standard of living in the face of falling real incomes, the (stabilizing) impact on aggregate demand can be substantial. Depending on which case describes more accurately the experience of the US and OECD countries, reigning in household debt bubbles comes at a low or high cost in terms of forgone output growth and different policies are required. This means understanding the determinants of household sector borrowing becomes crucial.

Assessing the explanatory power of the three hypotheses of Table 1 is important because the policy conclusions following from them are very different. The ECH and the ISH imply prioritizing the fight against income inequality as a measure to avoid unsustainable debt accumulation. This would involve measures like increasing minimum wages, strengthening the bargaining power of workers and other policies that support income growth in the lower end of the distribution while curbing top incomes. Such policy measures would not only tackle the root cause of the problem under the ECH but would also limit the negative growth effects of reigning in household debt accumulation because with lower income inequality debt would not be needed any more as an income substitute. Thus under the ECH and ISH simply restricting household borrowing without solving the underlying distributional problem could lead to substantially lower growth. On the other hand, if property price bubbles were the core mechanism leading to soaring household debt levels, restricting equity extraction and introducing loan to income ratios as benchmarks as well as increasing the supply of public housing in order to keep price rallies in check, would be more adequate policies to achieve and maintain a sustainable balance sheet structure of the household sector and thus economic stability.

Assessing the explanatory power of these hypotheses cannot be done in a single framework using a single data set. Different types of data have different strengths and limitations which make them more useful for some questions and less useful for others. The approach in this thesis is to rely first on US household survey data and then use an aggregate macro-panel covering 13 OECD countries.

Using household survey data is particularly helpful for investigating how the relative position within the households' peer group affects its borrowing behaviour. Therefore detailed household survey data will be especially useful for testing the ECH. However household survey data does not allow to directly assess the impact of interest rates and monetary policy as well as shifts in the credit supply. For these purposes aggregate data is better suited. The details about how these two different data sources are used are outlined at the beginning of each chapter.

The remaining thesis is structured as followed. Chapter 2 summarises the relevant theoretical and empirical literature. Chapter 3 provides an in-depth descriptive analysis of borrowing patterns of the US household sector based on the Survey of Consumer Finances (SCF). Seven stylized facts are presented and used as a benchmark against which theoretical models of household borrowing decisions are compared. The chapter concludes that real estate related borrowing (housing hypothesis) explains a large part of the increase in household liabilities over the sample period of 1989-2007. Shifts in credit supply conditions and borrowing due to negative income shocks (income stagnation hypothesis) also carry some explanatory power while the data does not support the notion that increasing income inequality (expenditure cascades hypothesis) was the key factor in explaining rising debt levels. This latter conclusion applies especially to the 2001-2007 period.

Chapter 4 investigates the factors driving US household borrowing up to 2007 in an econometric model. The focus of the chapter is on comparing the explanatory power of the expenditure cascades hypothesis and the housing hypothesis. In order to do so a method for obtaining individual household borrowing figures despite the lack of a panel structure from the Survey of Consumer Finances (SCF) is developed. This is the first time that the high quality information the SCF provides is used to investigate the impact of shifts in the income distribution and asset prices on household borrowing in an econometric model. The findings indicate that it is the interaction between the concentration of income at the top of the distribution and rising real estate prices which explains a large fraction of the increase in household borrowing prior to 2008. Therefore, neither the expenditure cascades hypothesis nor the housing hypothesis are, in isolation, sufficient in order to understand household debt accumulation.

Chapter 5 shifts from SCF data to aggregate macroeconomic data. Also the unit of analysis is changed, as a panel of 13 OECD countries over the period 1980-2007 is investigated instead of only the US. All three hypothesis are assessed and results indicate that real estate prices were important drivers of household borrowing and consumption spending and thus support the housing hypothesis. In contrast the data does neither support the expenditure cascades nor the income stagnation hypothesis as a general explanation of debt accumulation across OECD countries.

The final chapter 6 summarises this set of diverse results and draws some general conclusions as well as policy recommendations. Ways of addressing some of the shortcomings in future research are also pointed out.

2 Household Debt in Macroeconomic Theory

Household debt has not played a prominent role in most of macroeconomics. While it has long been a neglected issue, the effects of household debt have recently attracted interest, however there is comparatively less research on the determinants of household debt. Broadly speaking this thesis will distinguish between mainstream and heterodox approaches and approaches that derive debt from consumption decisions and those that derive it from asset transactions.

On the neoclassical side, in the textbook model of consumption (Romer 2012), the lack of household debt is due to the assumption of households maximising life-time utility, which implies smoothing consumption over the life cycle. In a growing economy, the young would save for retirement and the old would dis-save. If the income of the young is below their life-time consumption optimum (say, because they attend university), they would take out loans. Those in their prime age would save for retirement and the old would dis-save, using up their life-time savings. Debt would depend on demography and in the absence of demographic shifts should be stable in relation to aggregate income. Life-time consumers also attempt to smooth consumption in response to temporary and permanent shocks to life-time wealth (including discounted income streams). The impact of such shocks on household borrowing and consumption depends on the assumptions of the specific model one uses, in particular whether the model assumes credit constrained consumers and/or income uncertainty and buffer stock savings. Carroll (2001) argues that credit constraints and buffer stock saving behaviour is hard to distinguish empirically because in both cases households will strongly react not only to permanent but also to transitory income shocks. Because this thesis investigates the steady increase in household borrowing in OECD countries, temporary shocks are not of interest for this analysis. The impact of permanent shocks to income and wealth are discussed in the next two subsections. In the absence of such shocks the life-cycle model offers two other explanations for permanently increased household borrowing: Shifts in the age structure of the population (a declining proportion of young people reduces the need to borrow) and shifts in credit supply conditions (rising willingness to lend by financial institutions has the potential to ease liquidity constraints permanently).

Keynesian economics does not employ the life time optimising assumptions of neoclassical economics. Post Keynesian (PK) economics assumes given marginal propensities of consumption that will differ by class position or by the position within the income distribution. The rich save more. As the basic Keynesian model assumes autonomous consumption, this implies that if there are demand shocks, which there are plenty of in a Keynesian economy, parts of the population (namely the poor or those with illiquid assets) will have to accumulate debt to accommodate to the resulting income declines while autonomous consumption stays constant. However, household borrowing traditionally did not play a role in Keynesian modelling.[3] As soon as a gross wealth component is included in the consumption function one can argue that a positive wealth effect will be associated with a rise in household borrowing if taking on debt is used as a tool to 'realise' capital gains without selling the underlying asset. Thus, while consumer borrowing did not feature prominently in Keynesian macroeconomics until recently. Brown (1997) is an early exception who emphasizes the role of shifts in credit supply conditions for household borrowing. More recent studies in the Post Keynesian tradition which take borrowing of households into account are discussed in section 2.3.

The starting point and point of reference for the remainder of this section is the life-cycle model which predicts permanent increases in household borrowing only as a reaction to permanent wealth increases and assigns no role to the distribution of income. The second group of models reviewed in this section, emphasises the relative position of households within the income distribution in determining spending and borrowing decisions. While largely ignoring wealth effects, this literature argues that strong income growth at the top causes debt-financed spending sprees further down the distribution. Thirdly, models belonging to the Post Keynesian tradition are reviewed with a special focus on Minskian models. Adopting Minsky's ideas to the household sector provides a framework where wealth effects play the key role in explaining household indebtedness.

2.1 The Life-Cycle Model of Consumption

Textbook economic theory suggests households borrow in order to smooth out life-time consumption as a reaction to transitory income and wealth shocks (Romer 2012; Attanasio & Weber 2010). As the goal of this thesis is to explain the sustained increase in household debt to income ratios since the early 1990s, transitory shocks cannot be a valid explanation because borrowing as a reaction would only be transitory as well. In contrast, permanent income shocks in a life-cycle model of consumption imply permanent changes of consumption but do not affect borrowing. A permanent reduction in income will result in lower consumption expenditures as rational households would not take on debt as a substitute for income because higher debt levels would reduce future consumption or violate the budget constraint and thus the assumption of no default. Permanent shocks to wealth and most importantly housing wealth in contrast, can have a lasting effect on borrowing behaviour. If households were to consume some part of their real estate capital gains they would have to either sell the entire property or, more conveniently, take on an additional mortgage. Two important implications arise: First, the life-cycle model relies on rational consumers which can distinguish between transitory and permanent changes to income and wealth. Second, even permanent changes to income or wealth should only lead to small changes in consumption or

[3] This may be because of the role the government sector in general and unemployment insurance systems in particular play in how households accommodate to demand shocks without accumulating debt. Instead the public sector increases its deficit to stabilise the economy.

borrowing because these gains are consumed over the entire life-time of the household[4]. If the simple model outlined so far is extended by introducing credit constraints (Jappelli & Pagano 1994), constrained households would react to permanent increases of their housing assets by additional borrowing and thus less saving in the aggregate. These effects would be large even with long optimization horizons. Thus the life-cycle model predicts 'pure' wealth effects to be small due to long optimization horizons whereas wealth effects due to credit constraints might result in large changes of household consumption and borrowing. At the individual level permanent wealth shocks only boost consumption and borrowing of asset holders. Renters will scale back their expenditures and save more as a reaction to a permanent increase in property prices.

In contrast to permanent income changes, a change in the distribution of income has no impact on aggregate spending or borrowing because agents are assumed to be homogeneous in their preferences. Thus prolonged rises in household debt are only in line with the theoretical framework of the life-cycle model if they are driven by permanent increases in (housing) wealth. However as the financial crisis demonstrated not only households but also highly trained economists are likely to fail in distinguishing transitory and permanent changes properly. Alan Greenspan relied exactly on this underpinning when he argued that household debt is of no concern because it is justified by permanent increases in the value of households' assets (Greenspan 2004); only to be proven wrong by the financial crisis which followed.

The empirical literature which seeks to quantify consumption wealth effects within a life-cycle framework largely confirms the existence of such wealth effects (Lehnert 2004; Bostic et al. 2009; Cooper 2012; Salotti 2012; Juster et al. 2005; Dynan 2010). However since consumption wealth effects only provide indirect evidence for the hypothesis that house price appreciations drive household debt trends, this literature is only briefly discussed and interested readers are referred to the recent surveys by Cooper and Dynan (2016) and Paiella (2009). Interestingly, for most authors the distinction between transitory and permanent shocks is not of very high priority when it comes to assets. One of the core findings of this literature is the importance of credit constraints and thus large and positive housing wealth effects for credit constrained consumers. Bostic et al. (2009) are an exception, they do not find statistically significant wealth effects for credit-constrained consumers. Strategies to identify credit constrained households range from allowing wealth effects to differ between young and old consumers (Lehnert 2004; Bostic et al. 2009; Salotti 2012), relying on self-assessment (Bostic et al. 2009) and using liquid-wealth-to-income and debt-service-costs-to-income ratios (Cooper 2012).

[4] However as Carroll (2001) points out assuming (realistic) planning or optimizing horizons of two to three years instead of two to three decades results in much larger current effects of permanent changes.

All studies which directly assess the impact of asset prices on household borrowing report a positive direct relationship. Hurst and Stafford (2004) for example use PSID data and find that liquidity constrained households are more likely to refinance their mortgage and extract equity compared to unconstrained households. They use unemployment and low holdings of financial assets as indicators of liquidity constraints. Even though Hurst and Stafford (2004) focus on equity extraction as a reaction to transitory shocks they provide some evidence that house price appreciations are a potential factor which drives up household borrowing due to equity extraction. Haurin and Rosenthal (2006) are interested in the extent to which home equity extraction can explain consumer spending. Their regression of housing capital gains and other controls on household debt yields a positive and statistically positive effect based on data from the SCF and the National Longitudinal Survey of Youth. Thus Haurin and Rosenthal (2006) provide evidence of a positive link between house price appreciation and household borrowing. Dynan and Kohn (2007) have access to the unpublished version of the SCF and argue that house price appreciation was a major reason for rising household indebtedness. Finally Mian and Sufi (2011) analyse the variation across metropolitan statistical areas in household debt growth for homeowners between 2002 and 2006. By using a measure of housing supply elasticity as an instrument for house price growth they are the first to demonstrate that house prices matter independent of any potentially omitted third factors[5]. Mian and Sufi (2011) find large effects of house prices on borrowing especially for homeowners with low credit scores and high propensities to borrow on credit cards which they interpret as evidence of credit constraints and self-control problems, respectively.

Overall the empirical literature inspired by life-cycle models, supports the existence of wealth effects in general and the existence of credit constraints and thus (large) wealth effects for those households in particular. Thus there are strong reasons to believe that households’ assets should be taken into account when studying household borrowing behaviour. In contrast the life-cycle model does not assign any role to the distribution of income or the expenditure cascades hypothesis (Romer 2012, p.368).

[5] The argument is that asset prices, borrowing and consumption might be driven by a third factor omitted from the model. The two most important candidates which are likely to cause problems along these lines when estimating reduced form models are productivity shocks and credit supply shifts. Strategies to tackle that problem other than finding a relevant instrument for house prices include to split the sample between old and young households. If a common cause such as productivity drives asset prices and consumption, the increase in consumption should be most pronounced for young households since they benefit the longest from increased productivity. However since young homeowners are also more likely to be credit constrained this strategy is not ideal. Another approach is to split the sample between asset holders and non-asset holders. If there is a direct effect of increasing house prices on consumption and borrowing this effect should be positive and large for homeowners and small and negative for renters. However if a common factor is the driving force it should affect owners and renters alike. Such a sample split requires asset price measures at the regional level however. Mian and Sufi (2011) are the only paper working on the US case to deal with that issue. Other examples include Campbell and Cocco (2007) and Attanasio et al. (2009) using pseudo-panels based on UK data.

2.2 The Distribution of Income and Household Borrowing

A recent paper emphasises the importance of the relative position within the income distribution for household spending and borrowing decisions (Frank et al. 2014). Tracing that idea back to Veblen (1899) and Duesenberry (1949), Frank (1985) and Frank et al. (2014) argue that consumption is not only about satisfying one's needs but also about signalling and demonstrating social status ('keeping up with the Joneses'). They use that argument to explain household indebtedness: If status comparison manifests in an upward looking manner, rising income inequality will trigger a debt-financed consumption spree which cascades down the income distribution. The intuition is that households which experience slower income growth relative to their peers at the top of the distribution will start to borrow in order to keep up with the lifestyle of their peers. Since also debt-financed expenditures are relevant for status signalling, the process of borrowing in order to keep up with a life style of richer peers cascades down the income distribution.

The pioneering data work of Atkinson, Piketty and co-authors (Atkinson et al. 2011; Piketty 2014), revealed the sharp rise in income inequality over the last decades and contributed to the popularity of this idea. The result was several papers which explore the idea of expenditure cascades in formal models. Belabed et al. (2013) for example develop a multi sector stock-flow-consistent macro model in which households' consumption depends on richer peers' expenditures. Simulation results suggest that higher personal income inequality lead to debt-financed consumption cascades and current account deficits when financial markets are unregulated. Kapeller and Schütz (2014) combine the expenditure cascades idea with Minsky's idea of financial cycles in a stock-flow consistent framework. In their model households borrow as a reaction to increasing income inequality and eventually become over indebted. At that point banks will stop lending and the mass default of households triggers a crisis. Ryoo and Kim (2014) build a dynamic model which explains debt-financed consumption through emulation behaviour oriented towards the rich but do not ensure stock-flow concistency. Cardaci (2014) studies expenditure cascades in an agent-based stock-flow-consistent model and Nikiforos (2015) argues that expenditure cascades arise naturally out of US data if one takes financial balance constraints of the various economic sectors into account. In contrast Holt and Greenwood (2012) focus on the role of housing instead of consumption in the process of inter-household status comparisons. On a descriptive basis Holt and Greenwood (2012) argue that strongly growing top incomes led to an upward pressure on US house prices prior to the 2007 crisis. Since housing is an especially visible good and since in the US the neighbourhood is a primary determinant for the quality of many public services, most importantly schooling, many middle class households helped fuel the housing boom by pushing into high priced areas. Thus Holt and Greenwood (2012) raise the important issue that status signalling might happen mostly via housing and in such a scenario the emphasise on consumption of top income groups might be misleading.

There are several authors relying on aggregate data to test the expenditure cascades hypothesis. This literature is reviewed in detail in chapter 5. There is a high degree of heterogeneity with respect to the debt measures used and the specifications of the empirical models and thus direct comparisons are not always possible. Overall most studies find a positive impact on household borrowing (Behringer & Treeck 2013; Gu & Huang 2014; Klein 2015; Kumhof et al. 2012; Malinen 2014; Perugini et al. 2016) while Bordo & Meissner (2012) find a negative one. Importantly most of these studies do not control for fluctuations in asset prices which raises the potential for serious omitted variable bias problems.

A much richer literature uses micro data to investigate the impact of the position of an individual compared to its peers on that individual's behaviour. A closer look at this literature reveals a striking degree of heterogeneity in the way reference groups are defined. Many authors define peer groups based on household characteristics such as age and education (Maurer & Meier 2008), the region the household lives in (Ravina 2007), or just compare different categories of consumption goods (Heffetz 2011). These peer group definitions are fundamentally different from the upward looking status comparison which lies at the core of the expenditure cascades hypothesis. Nevertheless, all of these studies do find evidence of (at least) modest positive effects of peer group consumption on household spending. In addition, two recent studies explicitly adopt upward looking preferences and thus are important reference points for this paper. Bertrand and Morse (2013) model consumption for the bottom 80% of households depending not only on individual characteristics such as income or age but also on the consumption of the average household in the 80th percentile of the income distribution. Income percentiles are computed for each US State. Using data from the Consumer Expenditure Survey (CEX) from 1980 to 2008, they find a positive effect of consumption expenditures of rich households on lower income households and interpret these findings as evidence in favour of the expenditure cascades hypothesis. Carr and Jayadev (2015) model the relative position of the individual household in more detail. They use a measure of relative income defined as the proportion of households which are richer than household *i*. Using data from the Panel Study of Income Dynamics (PSID) from 1999 to 2009 Carr and Jayadev (2015) find positive and statistically significant effects of their relative income measure on household borrowing growth and interpreted these results as consistent with the expenditure cascades hypothesis. Coibion et al. (2016) use data from the Consumer Credit Panel and the SCF and they find that debt accumulation relative to income was lower for low income households living in a high income inequality area in 2001 compared to low inequality areas. Equivalently, high income households in high inequality areas borrowed more compared to high income households in low inequality areas.

Overall the empirical evidence seems to weakly support the expenditure cascades hypothesis. This is because in contrast to studies that use microeconomic data, studies which rely on aggregate data are not conclusive about the existence (Behringer & Treeck 2013; Perugini et al. 2016) or non-existence (Bordo & Meissner 2012; Stockhammer & Wildauer 2016) of a positive link between income inequality and (household) borrowing. Also as pointed out by Bertrand and Morse (2013) the existence of a positive link between relative income and household debt is not sufficient for demonstrating that the expenditure cascades hypothesis is able to explain a major proportion of the increase in household debt prior to 2007. Bertrand and Morse (2013) report that based on their results the increase in income inequality from the 1980s to the mid-2000s only explains a 1.3 percentage points out of a 9 percentage point fall in the NIPA savings rate. Therefore rising income inequality cannot be regarded as the main cause for increased household indebtedness. In addition despite strong evidence of the important role housing wealth plays for household spending and borrowing decisions, the studies seeking to test the expenditure cascades hypothesis as an explanation for rising household debt, generally ignore the role of assets. Another issue with survey based papers is that both Bertrand and Morse (2013) as well as Carr and Jayadev (2015) use data which only provides limited detail about the top of the income distribution. In particular there is no study on the expenditure cascades hypothesis using SCF data. This is an important drawback of existing studies since income inequality in the US rose due to strongly growing top incomes and thus not including those top households in one's sample will yield misleading results. The next section elaborates on that latter point by discussing the specific advantages the SCF provides for studying the links between household borrowing, asset prices and the distribution of income.

An explanation of increased household borrowing very similar to the expenditure cascades hypothesis rests on the observation that households do not want to cut consumption below levels already reached in the past or below a certain minimum level. For example Barba and Pivetti (2009) argue that rising inequality has pushed poor households into debt as they were trying to maintain their standard of living. Kumhof and Ranciere (2010) adopt the latter approach and show that a decline in bargaining power of workers leads to increased income inequality and results in a debt-financed attempt to maintain living standards. Brown (2007) argues that prior to the 2008 crisis US households relied on borrowing as a substitute for slow income growth. Despite the fact that the exact mechanisms differ, these models will be regarded as variants of the income stagnation hypothesis. It is another way of linking rising income inequality to rising debt levels however without relying on status comparison as the underlying driving factor.

With respect to the distribution of income Post Keynesian economic theory assumes different propensities to consume across the income distribution due to empirically higher saving rates towards the top of the distribution (Dynan et al. 2004). In models without a personal distribution of income a higher propensity to consume out of wages than out of profits is assumed. These assumptions imply that a relative increase of top or profit incomes will lead to a decrease in aggregate consumption spending (Hein & Vogel 2008; Onaran et al. 2011). A priori it is not clear whether such a redistribution would also increase household borrowing. It would do so if households want to sustain past consumption levels and are not liquidity constrained.

2.3 Minsky and Stock Flow Consistent Models

With the collapse of the US housing market beginning in 2006 and the following financial crisis the Post Keynesian economist Hyman P. Minsky and his Financial Instability Hypothesis (FIH) (Minsky 1978) have attracted renewed interest. The core idea of the Financial Instability Hypothesis is to reckognize the importance of business finance. Firms, take out loans to run their daily business (working capital) and to invest in productive assets. While businesses differ with respect to the riskiness of how they finance their activities, the FIH states that during an economic boom firms generally will grow more comfortable and start to switch to riskier financing structures. The reason for such a switch can be due to increasing asset prices their balance sheets improve or due to growing optimism aobut the future business environment. This mechanism is summarised by the powerful notion that periods of stability breed instability. Instability emerges from more and more firms engaging in riskier financing structures and eventually a trigger event (such as interest rate increases by the central bank) will lead to a quickly growing number of corporate defaults. In order to talk about financial stability, Minsky distinguishes three broad financing schemes of firms. First, he uses the term save financing in order to describe firms which are able to meet interest payments as well as principal repayments out of their current cash flow. Second, risky finance describes firms which are able to make interest repayments out of their operating cash but need to roll over their liabilities in order to make principal repayments. Third, speculative finance describes a state in which firms' cashflows are not even sufficient to make interest payments and firms have to rely on new borrowing for interest payments as well as principal repayment. Minsky's framework was briefly mentioned as a theory able to explain major crises in contrast to standard New-Keynesian or Real Business Cycle models. Minsky's work is relevant not only in the context of firms but also in the context of household debt accumulation because it provides an explanation of why agents take on too much debt and eventually default. In analogy to Minsky's original work Ryoo (2015) is currently the only paper which applies Minsky's ideas to the household sector and develops a model of endogenous instability. Kapeller and Schütz (2014) and Ryoo and Kim (2014) also adopt Minskyian ideas to the household sector but combine that with the expenditure cascades hypothesis and thus these authors have been discussed in the previous subsection. In Ryoo (2015) households borrow in order to finance additional consumption and housing expenditures. Increasing housing demand leads to capital gains which feeds back into additional borrowing due to the improved balance sheet position of households. The result is a prolonged boom followed by a severe recession. The key point is that even if rising house prices are driven by a growing debt-load and thus cannot be permanent, they encourage household borrowing. Thus a Minskyian model of the household sector does not depend on the questionable distinction between transitory and permanent shocks but predicts positive wealth effects on consumption and offers a powerful tool for understanding households' debt accumulation prior to 2007. While Ryoo (2010, 2013) explicitly models household and firm balance sheets in an SFC framework, it is still firms that hold the debt. This highlights the fact that Minsky's *Financial Instability Hypothesis* is a theory about the behaviour of the corporate sector. As a result most formal models of that idea only deal with corporate debt. For recent examples of corporate Minsky models see Bhattacharya et al. (2015), Keen (2013), Charles (2008), Dos Santos (2005) and the references therein.

A striking weakness of the Minskyian tradition is the lack of empirical papers investigating the interaction of assets, household debt and consumption, even among Post Keynesian authors, which only began to change after the crisis. Borio (2014) cites Minsky and argues that the financial cycle which he describes as "self-reinforcing interactions between perceptions of value (…) and financing constraints, which translate into booms followed by busts" can be reasonably described by the interaction of household debt and property prices. Bezemer and Zhang (2014) establish a positive link between house prices and household debt but also find that credit booms with high shares of mortgage credit in total bank credit are more likely to go bust in a panel of 37 countries. Closely related is also the work of Schularik and Taylor (2012) who find a positive link between credit and the likelihood of financial crises but they look at private bank credit and are not especially interested in the household sector. Bezemer et al. (2014) analyse lending decisions of banks and highlight the distinction between nonfinancial credit and asset market credit. The latter includes credit to financial institutions and mortgages. Bezemer et al. (2014) then argue that the two types of credit will have different demand and growth effects. Their research is on the effects of household debt rather than its causes, but it is useful here as it highlights that households engage in asset transactions and therefore a substantial part of household debt is not related to consumption. Arestis and González (2013) propose a model of the supply and demand for housing. The model allows for private sector borrowing derived from real estate transactions, however it does not explicitly model business and household borrowing decisions separately.

Beyond the Financial Instability Hypothesis, Post Keynesian economic theory provides another rationale for the role of asset prices in household borrowing. The tradition of stock-flow-consistent (SFC) modelling, going back to Tobin (1969; 1982) and pioneered by Godley and Lavoie (2007) emphasizes the importance of connecting all flows in an economic model to corresponding stocks at the end of the period. An important implication of these models is that economic sectors are thought of in a balance sheet context and assets and liabilities will enter behavioural equations. This leads to the notion that economic actors are anchored by so-called stock-flow norms (Godley & Lavoie 2007, p.75). In the case of households this means they form consumption decisions in line with achieving a desired net-wealth-to-income ratio. These desired wealth-to-income ratios reflect propensities to save and consume out of disposable income and net-wealth respectively and thus are influenced by precautionary saving motives as well as social norms. The implication is that if household sector (real estate) wealth increases up to the point where it surpasses the target ratio, households will consume that 'excess wealth'. Thus since housing wealth is the most important asset for households, borrowing in order to 'realise' real estate capital gains would be the result for asset holders. Zezza (2008) develops a rich stock-flow-consistent model where he explicitly includes stock and real estate markets. His consumption function exhibits a wealth effect. Thus also if one abstracts from the role assets play as collateral, Post Keynesian theory predicts a wealth effect positively impacting consumption and household borrowing due to the adjustment towards desired wealth-to-income ratios.

To sum up, Post Keynesian theory strongly predicts the existence of wealth effects. First, due to the role assets play as collateral and their ability to ease credit constraints. Second, households will increase spending and potentially borrowing as a reaction to rising asset values because desired wealth-to-income ratios, so-called stock-flow-norms, are reached and require less saving efforts.

2.4 Reconciling the Three Hypotheses with Macroeconomic Theory

This section lays out the connections between the three central hypotheses (ECH, HH, ISH) and the different theoretical approaches to economics discussed in the previous sections. The expenditure cascades hypothesis rests on the idea of other regarding social norms often used by authors in the behavioural economics tradition. Eckerstorfer and Wendner (2013) for example study how other-regarding social norms affect consumption (they use the term consumption externalities) and derive optimal tax rates. Frank et al. (2014) study the evolution of saving rates which they argue are strongly influenced by other-regarding social norms. However also outside behavioural economics, researchers have embraced this idea. There are for example several recent contributions in the Post Keynesian tradition (Kapeller & Schütz 2014; Belabed et al. 2013; Ryoo & Kim 2014) which model consumer behaviour dependent not only on own income but also on the consumption of a peer group. So while other-regarding social norms in general and the expenditure cascades hypothesis in particular can be incorporated into different types of models, they are not compatible with the textbook version of the life-cycle model of consumer behaviour as pointed out by Romer (2012, p.368) because rational consumers would realize that borrowing for current consumption would reduce future consumption.

The housing hypothesis is consistent with several ideas expressed by different schools of thought. First, if households' perceptions about the future are determined by the value of their assets, then rising asset prices lead to optimistic and eventually over-confident consumers which take on large amounts of debt. This is a Minskian interpretation of the housing hypothesis because in Minsky's

Financial Instability Hypothesis growing confidence and eventually overconfidence due to rising asset prices and prolongued periods of strong economic growth encourages firms to rely on increasingly risky financing structures which eventually leads to mass defaults and a crisis. Second, if households are credit constrained then rising property prices will ease these constraints at least for those households that own property. As a result, their borrowing will go up, a behaviour consistent with a neoclassical life-cycle model incorporating credit constraints. Third, if households' consumption and saving decisions are anchored by stock-flow norms as argued by authors in the stock-flow-consistent modelling tradition, then an increase in asset prices will raise the current wealth to income ratio above the target ratio and households will start to consume the excess wealth. Some of them will borrow in order to avoid the selling of illiquid assets.

The income substitution hypothesis is based on self-regarding social norms. Like the ECH this idea can be incorporated in different theoretical frameworks. Kumhof and Rancière (2010) are an example of incorporating it into a neoclassical DSGE model. However one of the earliest proponents of the idea in economics, James Duesenberry (1949) can be regarded as a Keynesian economist. So as with the ECH the ICH represents a specific behavioural assumption which can be incorporated into different frameworks. It is important to note however that the idea of self-regarding social norms is not compatible with the textbook version of the life-cycle model and also incorporating liquidity constraints into the latter do not make it compatible with self-regarding social norms. The reason is that even if households lose income relative to the past, if they are rational as the lifec-cycle model assumes, they would not borrow in order to make up for that income loss, as the resulting interest payments would reduce their future consumption.

In addition to the three main hypothesis the analysis will allow to assess the importance of the age structure of households. This is important because the life-cycle hypothesis in its textbook but also in extended versions (for example including credit constraints), predicts a major impact of household age and especially shifts in the age structure on household borrowing. However the data does not support this prediction as the coefficients on household age are not statistically significant in the analysis undertaken in chapter 4 and the sign on the old-age dependency ratio used for the analysis in chapter 5 suggests that borrowing declines as the population becomes older, which is at odds with the life-cycle hypothesis.

This brief discussion highlights that there is no 1:1 relationship between the tested hypotheses and competing theoretical approaches to economics. However, since the goal is not to test entire theories but only hypotheses this is not a reason for concern. While testing different hypotheses is less far-reaching than testing entire theories it still allows to identify key factors and potentially groups of models within several theories which are more useful and more consistent with the data than others.

3 A descriptive Analysis of US Household Debt

This chapter presents a descriptive analysis of the US' household sector borrowing patterns and discusses to what extent the observed data is consistent with the three hypotheses of interest. In particular the chapter addresses three questions: First, to what extent did increases in household debt coincide with polarizations in the distribution of income? A failure of such a co-movement between debt and income inequality would be inconsistent with the ECH. Second, did household debt increases coincide with increasing real estate wealth and real estate prices? Observing a direct relationship between these two variables would be consistent with the housing hypothesis. Third, to what extent can the borrowing behaviour of households with stagnant incomes explain aggregate trends? If most household borrowing prior to the crisis was a reaction to stagnant income growth, this would support the role of the ISH in explaining rising US debt levels. In addition the chapter also sheds light on the impact shifts in credit supply conditions had on household borrowing decisions.

The findings can be summarised as follows. First, US households increased their liabilities substantially. Between 1989 and 2007 total outstanding liabilities of the household sector rose from \$4.4 trillion to \$12.7 trillion in real terms[6]. This increase was the result of two effects: households taking on more debt as well as more households taking on debt. Second, the data strongly points to the dominant role of the housing market as a driving factor of household debt accumulation. More than 80% of outstanding household liabilities are mortgages and most of them were taken out in order to purchase or improve property. Third, the nexus between intra group income inequality and household indebtedness is shaky. The data supports such a relationship for black households but not for white households in the pre-crisis period between 2001 and 2007. Fourth, changes in credit supply conditions explain some of the increase of household debt especially for the group of white college-educated households. Fifth, the notion that borrowing was concentrated among households whose incomes were stagnant and which used debt to maintain living standards, is not confirmed by the data.

The next section briefly introduces the specific characteristics of the Survey of Consumer Finances (SCF). The main part consists of seven sections discussing the data and a final section concludes.

3.1 The Survey of Consumer Finances

3.1.1 The SCF and Non-Observation and Non-Response Bias in Surveys

This thesis relies on SCF data between 1989 and 2013. The SCF is a tri-annual survey undertaken by the Federal Reserve Board (FRB) in cooperation with the Statistics of Income Division (SOI) of the Internal Revenue Service (IRS). In each wave between 3,143 (1989) and 6,482 (2010) observations are included. The SCF focuses on household income, assets and liabilities and represents the most

[6] Throughout the chapter monetary values are expressed in 2013 prices if not stated otherwise.

detailed source of information about household balance sheets and especially high income household balance sheets. This latter benefit of the SCF stems from the dual-frame sample design consisting of an area-probability sample and a list sample.

About two thirds of the completed cases stem from the area probability sample which is built in three stages. In the first stage metropolitan areas and rural counties are stratified by a variety of characteristics and primary sampling units (PSUs) are selected proportional to their population. At the second stage subareas are selected within PSUs again based on stratification. At the third stage random samples are drawn within these subgroups. This design ensures that each household in the sample has the same probability of being selected. Thus the area-probability sample covers broadly distributed household characteristics, while at the same time limiting the cost of data collection due to the stratified design.

However there are two important shortcomings. First, due to the highly skewed distribution of household characteristics like income and wealth, an enormous sample size would be needed to gain sufficient observations of wealthy households to obtain an adequate picture of the distribution of these characteristics. The cost of obtaining such a sample would be substantial. For example in 2012 the relatively large Consumer Expenditure Survey (CEX) is based on less than 14,000 observations representing 124 million households. This corresponds to a sample of only 0.11‰ of the target population. Even a large survey like the CEX is not sufficient to adequately represent the highly skewed income distribution because not enough observations from the top end of the distribution are part of the sample. This fact becomes obvious if one compares the average pre-tax income in the 10th decile calculated from CEX data with data published by the IRS. According to the IRS, average income in the top decile in 2011 was $2.1 million[7] compared to $229,771[8] in 2014 according to CEX data. So despite the timing gap and different income concepts, the difference is striking. This is referred to as the problem of non-observation which is demonstrated by simulation exercises in Kennickell (2005) and Eckerstorfer et al. (2016).

Second, there is evidence that the likelihood of participation in (wealth) surveys is negatively correlated with household wealth itself (Kennickell & McManus 1993; Singer 2006; Kennickell 2008), known as systematic non-response. One can think of several reasons why rich households are less willing to participate, ranging from greater concerns about data protection to higher valuation of the time needed to complete the interview. However if non-response is systematically related to household characteristics like wealth or income, any estimates based on samples which do not address this problem will be biased.

---

[7] IRS data based on “1979 Income Concept” from SOI Bulletin article - Individual Income Tax Rates and Tax Shares, Table 7.
[8] Summary Table 1110 (http://www.bls.gov/cex/2014/combined/decile.pdf).

In order to deal with these two problems, the SCF relies on the second component of the dual-frame sample: the list sample. The purpose of the list component is to over-sample wealthy households. In order to be able to identify such households prior to data collection an external data source is needed. Due to a cooperation under extremely strict privacy conditions, the list sample is built by using a sample of income tax records. Based on that information the assets of tax units are estimated by capitalising asset related income streams[9]. Then observations are selected in a two stage process. First, only observations in PSUs which were already selected for the area-probability sample are considered in order to keep the costs of the survey in check. Second, households are stratified based on percentiles of the estimated asset holdings. Then samples are drawn from each strata and strata corresponding to higher estimated net wealth are sampled at higher rates. The details about the sample design of the list sample are not publicly available in order to ensure anonymity of all participants.

Due to non-observation and non-response problems, surveys which do not pay as much attention to their sample design as the SCF does and in particular surveys which do not apply oversampling techniques suffer from serious shortcomings and are in general not able to provide an adequate picture of the income or wealth distribution. Kennickell (2008) demonstrates the importance of the list sample for the SCF: While net worth at the 90th percentile only increases by 5.5% due to the additional information from the list sample, at the 99th percentile the increase is 74%. Vermeulen (2014) and Eckerstorfer et al. (2016) demonstrate the impact of such a shortcoming for other countries and surveys. The latter paper estimates that aggregate net wealth is underestimated by about one quarter due to non-observation and non-response problems in the case of Austria in the Household Finance and Consumption Survey. Since the aim of this thesis is to assess to what extent rising income inequality can explain the increase in household debt, taking non-observation and non-response problems seriously is crucial.

Like any household survey the SCF faces the problem that participants are unable or unwilling to answer some of the questions. Leaving these questions unanswered would effectively lead to the loss of all information provided by that household (complete case analysis). Instead the SCF imputes missing values based on statistical modelling, referred to as multiple imputation (Rubin 1987; Kennickell 1998). In order to reflect the uncertainty associated with statistical modelling of missing information the process is repeated 5 times, yielding 5 separate datasets, so called *implicates*. Working with SCF data requires to analyse each dataset and then combine the results based on *Rubin's rules* which in the simplest form state that point estimates should be averaged across implicates and in order to obtain proper standard errors for these estimates one has to take into

[9] Kennickell (2000) provides a detailed discussion of the details and the two different models used.

account the variation within implicates as well as between implicates. For a more detailed discussion of these issues see section 3.1.3.

3.1.2 The SCF Compared to Other US Datasets

The fact that the SCF is seldom used for econometric modelling justifies briefly pointing out the most important differences when compared to other US household surveys notably the Consumer Expenditure Survey (CEX) and the Panel Study of Income Dynamics (PSID). First, the SCF and CEX are repeated cross sections[10]. While the CEX re-interviews households within the year, the SCF in general does not interview the same households twice except in 2009 when the 2007 sample was re-interviewed. In addition the SCF collects data in a three year interval while the CEX does so at quarterly and weekly intervals. In contrast the PSID is a panel which started collecting annual information in 1968. From 1999 onwards households are interviewed every other year. The PSID was initially designed to study the dynamics of income and poverty and thus started out by oversampling low income families. While the SCF and CEX rely on new samples on an annual and triannual basis respectively, the PSID sample only changes due to births, deaths and marriages occurring in the families originally sampled in 1968. In addition households stop participating for various reasons, which is known as attrition. In 1997 a major sample adjustment took place in order to better take into account immigration which occurred since the late 1960s. Second, the SCF focuses on income, assets and liabilities but contains almost no information on expenditures. In contrast, the CEX focuses heavily on expenditures while the PSID focuses on low-income, poverty and health. It becomes clear that all three data sources were designed for very different purposes.

Based on this brief comparison, three key advantages of the SCF are identified. First, using tax information to oversample wealthy households provides detailed information about the top tail of the income and wealth distribution. Second, by not spending interview time on household expenditures, the SCF is able to provide detailed information about households' balance sheets instead. Even though the CEX and PSID also collect information on assets and liabilities the SCF does it in a much more detailed way. Third, using a new sample for each wave minimizes the problem of attrition. Panel studies in general face the problem that if households drop out of the sample in a non-random way, the sample becomes unrepresentative over time. By drawing a new sample for each wave, repeated cross sections avoid this problem.

These highly attractive features of the SCF come at a cost however. At least two disadvantages are obvious. The first and most important drawback of the SCF is the low frequency at which data is collected. Three year intervals are long periods of time and by definition the SCF only provides limited information about what is going on in between. It is unthinkable to only have reliable information about inflation or GDP growth every three years. The second drawback is the design as

[10] Technically the CEX is a rotating panel but it follows households only over five quarters.

a repeated cross section which prohibits researchers from using panel data methods. Thus SCF data does not allow to control for (time-invariant) heterogeneity and also the dynamics of the household balance sheet cannot be studied in full detail at the individual level. With respect to the last issue, the next chapter demonstrates however that by using the rich information the SCF provides about the timing of borrowing decision that this problem can be solved at least partially (see section 4.2). Third, the SCF does not contain information on household expenditures which would be necessary in order to investigate wealth effects on consumption spending. Due to its focus on income and the household balance sheet however, this is a fundamental constraint due to the limited amount of time interviewers have when collecting information.

Despite these drawbacks, when it comes to studying household indebtedness and income inequality the SCF's advantages outweigh its limitations. Much better coverage of the upper tail of the income and wealth distribution and much more detailed information on household assets and liabilities when compared to other household surveys are strong reasons to rely on the SCF when investigating the links between household debt, real estate wealth and income inequality.

3.1.3 Some Special Characteristics of Survey of Consumer Finances Data

The Survey of Consumer Finances exhibits two special features which deserve comment as they heavily impact on how the dataset must be used to carry out any sort of analysis. The first of these is that the SCF comes as a multiply imputed data set as mentioned above. What this means is that in order to deal with missing answers on individual interview questions, the team at the Board of Governors of the Federal Reserve System which compiles the publicly available data relies on statistical modelling to infer plausible answers on these missing questions (Rubin 1987; Kennickell 1998). Rubin (1987) is considered to be the father of multiple imputation. The motivation for estimating missing values (imputing missing values) is that without it researchers would be required to stick to a complete case analysis and a lot of information would be lost. By using the information an interviewee provided on other questions to estimate missing answers, also partially incomplete interviews can be used for analysis. In order to reflect the uncertainty of these estimated answers five values are drawn from the estimated conditional distribution of the data (thus multiple imputation), yielding five values and ultimately five different datasets. The different datasets are called *implicates*. Properly using SCF data for quantitative analysis requires to carry out any data manipulation, any estimation and any computation of aggregates separately for the five datasets. The five final results are then averaged across all implicates. In the context of OLS and the computation of regression coefficients this means any regression must be carried out five times, once for each implicate, and then the point estimates are averaged. (Variance estimation is discussed below). The averaged coefficient is the value which is reported in the regression tables in chapter 4. The principle also applies to simpler statistics such as computing unconditional averages of variables or medians as in chapter 3. These statistics are computed for each implicated and then averaged across implicates. When the data is used in Stata format, the number of data points the dataset includes is six times the number of actual observations (interviews). The reason is that Stata also requires and saves the original version of the dataset without any imputation. Whenever I refer to observations I refer to the number of actual interviews whereas I use the term datapoints to describe the total number of datapoints stored in the dataset which is six times the number of actual interviews. Within Stata the family of commands which is required to deal with multiply imputed datasets is the “mi” family. An overview can be obtained by typing “help mi”. The code used to carry out the analysis discussed in chapters 3 to 5 is presented in Appendices V to VIII.

The second key feature researchers need to be aware of especially for hypothesis testing is that of complex survey designs. In the context of the SCF this means that the data is not obtained from a pure random sample. As outlined in section 3.1.1, randomly picking households from an address list would lead to two key problems. First, data collection would be very costly as individual households would be dispersed over the entire country and distances from one randomly selected household to the next closest could be large. Second, if certain characteristics are heavily concentrated among the population, random sampling will at best produce a very inefficient estimate (meaning large standard errors) of the population value of this characteristic. This second problem was referred to as non-observation bias in section 3.1.1. In order to tackle both issues the SCF is based on two different sample parts. The first is a multi-stage area-probability sample which is designed to cover broadly distributed characteristics while minimizing costs of collecting the data. About two thirds of the completed cases stem from the area probability sample which is built in three stages. In the first stage metropolitan areas and rural counties are stratified by a variety of characteristics and primary sampling units (PSUs) are selected proportional to their population. At the second stage subareas are selected within PSUs again based on stratification. At the third stage random samples are drawn within these subgroups. The second component of the SCF’s sample is called a list sample. The purpose of the list component is to over-sample wealthy households. In order to be able to identify such households prior to data collection an external data source is needed. Due to a cooperation under extremely strict privacy conditions, the list sample is built by using a sample of income tax records. Based on that information the assets of tax units are estimated by capitalising asset related income streams[11]. Then observations are selected in a two stage process. First, only observations in PSUs which were already selected for the area-probability sample are considered in order to keep the costs of the survey in check. Second, households are stratified based on percentiles of the estimated asset holdings. Then samples are drawn from each strata and strata corresponding to higher estimated net wealth are sampled at higher rates.

[11] Kennickell (2000) provides a detailed discussion of the details and the two different models used.

This description of the sampling process shows that standard ways of variance estimation which are based on a random sampling assumption do not apply. This means that standard errors and t-statistics reported by software packages when running standard estimation commands are not valid for SCF data. The reason is that these commands usually assume the data comes from a random sample. Variance estimation with the SCF requires to take both the imputation of missing data into account as well as the non-random aspects of the sample design. So based on Rubin's (1987) first rule the point estimate of a parameter of interest (β) is the average across all five implicates:

$$\hat{\beta} = \bar{\beta} = \frac{1}{5}\sum_{m=1}^{5} \hat{\beta}_m \tag{1}$$

Based on Rubin's (1987) second rule the variance of that estimator consists of the within implicate (W) and the between (B) implicate variance:

$$\widehat{Var(\beta)} = W + 6/5B \tag{2}$$

where:

$$W = \frac{1}{5}\sum_{m=1}^{5} \widehat{Var(\beta_m)} \tag{3}$$

$$B = \frac{1}{4}\sum_{m=1}^{5} \left(\hat{\beta}_m - \bar{\beta}\right)^2 \tag{4}$$

While the computation of the between variance (B), defined as the variance of the estimate across implicates, the computation of the within variance (W) requires knowledge of the sample process. If the data were coming from a pure random sample than standard procedures and formulas could be used. However this is not the case for the SCF and even worse, the details of the sampling process such as identifiers of primary and secondary sampling units, observations coming from the area probability or the list sample etc. are not published for the SCF. Instead the Fed's Board of Governors publishes a set of 999 replicate weights which can be used in a bootstrap procedure and which simulates the actual sampling process. Thus, $\widehat{Var(\beta_m)}$ is defined as:

$$\widehat{Var(\beta_m)} = \frac{1}{998}\sum_{w=1}^{999} \left(\hat{\beta}_{m,w} - \bar{\beta}_{m,w}\right)^2 \tag{5}$$

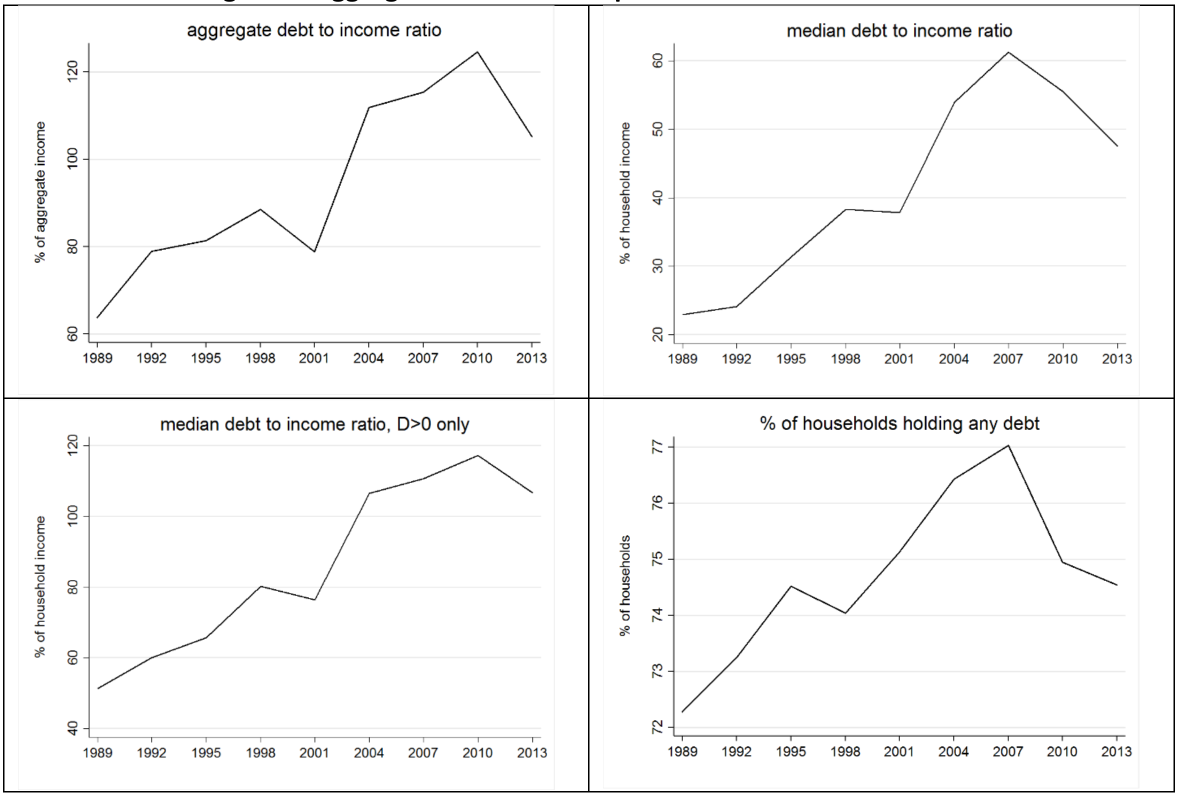
where $\hat{\beta}_{m,w}$ refers to the parameter estimated based on implicate *m* and replicate weight set *w* and $\bar{\beta}_{m,w}$ is the average parameter estimate across all replicate weight sets on implicate m. Taken together this means that running a single OLS regression in Stata and using the correct set of commands both with respect to the multiply imputed nature of the dataset and the fact that variance estimation need to be carried out by means of bootstrap, implies that the software carries out 5000 regressions in the background. The actual code used to carry out the analysis presented in this thesis is presented in Appendices V to VIII.

3.2 Debt to Income Ratios between 1989 and 2007

The first step of the descriptive analysis is to analyse by how much household debt increased prior to the financial crisis. Since the thesis aims to assess the validity of different explanations of growing household debt, clarifying by how much household debt actually increased is a natural starting point. The general finding is that US households took on large amounts of debt in the years leading up to the financial crisis of 2007/2008. The concrete numbers based on the SCF are as follows: Household disposable income as well as debt peaked in 2007 at $11 trillion and $12.6 trillion respectively (see Figure 1 in Appendix I), yielding an aggregate debt-to-incom ratio of 115%. The upper left panel of Figure 1 displays these ratios of aggregate outstanding household liabilities to aggregate disposable household income for all years for which SCF data is available. The total amount of US households' liabilities expressed as a percentage of total disposable income rose from around 60% in 1989 to 115% in 2007 and more than 120% in 2010. The peak in 2010 was the result of disposable incomes falling faster than liabilities in the severe recession which followed the financial crisis. These numbers are widely known and several economists stressed the importance of household debt and private sector debt in general as an informative indicator for financial crises (Schularick & Taylor 2012; Bezemer & Zhang 2014). The aggregate ratio however does not contain any information about how debt and income are distributed among households.

If the individual household debt to income ratios are very different across households then the aggregate ratio of household debt to income will not provide a good idea about the indebtedness of "the average" household. For this reason the upper right panel of Figure 1 displays the median debt to income ratio which is computed as the median of individual debt to income ratios. The striking difference compared to the aggregate ratio is that the median ratio is substantially lower and was only 23% in 1989. The first important reason why median debt to income ratios were so low is that over the sample period between 22% and 28% of US households did not hold any debt, which pulled the median closer to 0. The lower left panel of Figure 1 displays the median debt to income ratio if only indebted households are taken into account (i.e. zero debt to income ratios are excluded). Logically excluding non-debtors leads to higher median ratios. Nevertheless they are still substantially lower compared to the aggregate ratio, which points to a second reason for the difference between median and aggregate ratios: A large proportion of indebted households had quite low debt to income ratios and at the same time this group contained disproportionately low income households which means their contribution to aggregate income as well as debt measures was below average. The flip side of that statement is that these low median debt to income ratios indicate that a few heavily indebted households (relative to income as well as in absolute terms) dominate aggregate outcomes.

Figure 1: Aggregate household disposable income and liabilities



The aggregate debt to income ratio is computed as aggregate liabilities over aggregate income. The median debt to income ratios are the median of the individual household debt to income ratios. Source: author's calculations based on SCF.

Median debt to income ratios increased substantially over the sample period. In fact the median ratio in the upper right panel of Figure 1 almost tripled from 23% in 1989 to its peak at 61% in 2007. One reason for this sharp increase is that the proportion of indebted households grew strongly over the sample period, as can be seen from the lower right hand panel of Figure 1. This increase in the proportion of debtors pushed up median debt to income ratios as more people took on debt. Another reason for rising median debt to income ratios is that indebted households took on more and more debt, as can be seen from the lower left panel of Figure 1. So there are two factors driving median debt to income ratios in the upper right panel of Figure 1: more people took on debt (the proportion of debtors in the total population rose) and people took on more debt (the degree of indebtedness increased).

However a closer look at the median ratio including only debtors, reveals that there was a change in the debt accumulation dynamic after 1998. Between 1998 and 2007 the median ratio (including only debtors) and the aggregate ratio converged. This means that the concentration of liabilities in the hands of a relatively small number of heavily indebted households was easing up. The distribution of debt holdings flattened out: Table 1 presents how the relative weight of the 5% most heavily indebted households changed between 1992 and 2004. The measure of indebtedness used is the debt to income ratio. The first result from Table 1 is that debt was concentrated in the hands of a few heavily indebted households because only 5% owed almost 20% of all outstanding liabilities in 1992. This proportion however fell from 19% to 17% in 2004 in a shift which indicates that the concentration of debt eased up. The other side of the coin is that the share of total debt of those households which have had debt to income ratios around the median and were above the 25th percentile and below the 75th percentile of debt to income ratios, increased from 26% in 1992 to 32% in 2004.

Table 2: Concentration of debt with highly indebted households

percentiles of debt to income ratio distribution	year	share of total liabilities	total amount (2013 Dollars)
96-100 (Top 5%)	1992	19.1%	$0.9 trillion
96-100 (Top 5%)	2004	17.3%	$1.89 trillion
26-75 (Middle 50%)	1992	25.8%	$1.27 trillion
26-75 (Middle 50%)	2004	32.1%	$3.51 trillion

Percentiles correspond to the distribution of debt to income ratios not to the distribution of income. Source: author’s calculation based on SCF waves 1992 and 2004.

Several important conclusions can be drawn from the data presented in this section. First, US private households leveraged up considerably. This result holds regardless whether one looks at aggregate or individual debt to income ratios. Second, the increased debt holdings were the sum of two effects: more people taking on debt (as indicated by the increasing proportion of indebted households) and people taking on more debt (indicated by the increasing debt to income ratios of indebted households). Third, debt holdings in the household sector were highly concentrated in the hands of a few highly indebted households. The 5% of most indebted households (based on debt to income ratios) accounted for almost 20% of all outstanding liabilities in 1992 and 17% in 2004. Fourth, the concentration of debt in the hands of few highly indebted households eased up after 1998. The importance of the most heavily indebted households declined as the share of debtors in the population increased and the middle of the distribution took on more debt. Overall it seems that household debt accumulation did not only accelerate after 1998 but that the process also changed in a qualitative way with a broader part of the population taking on debt. This makes the immediate pre crisis period 2001-2007 of special importance.

3.3 Household Debt by Type

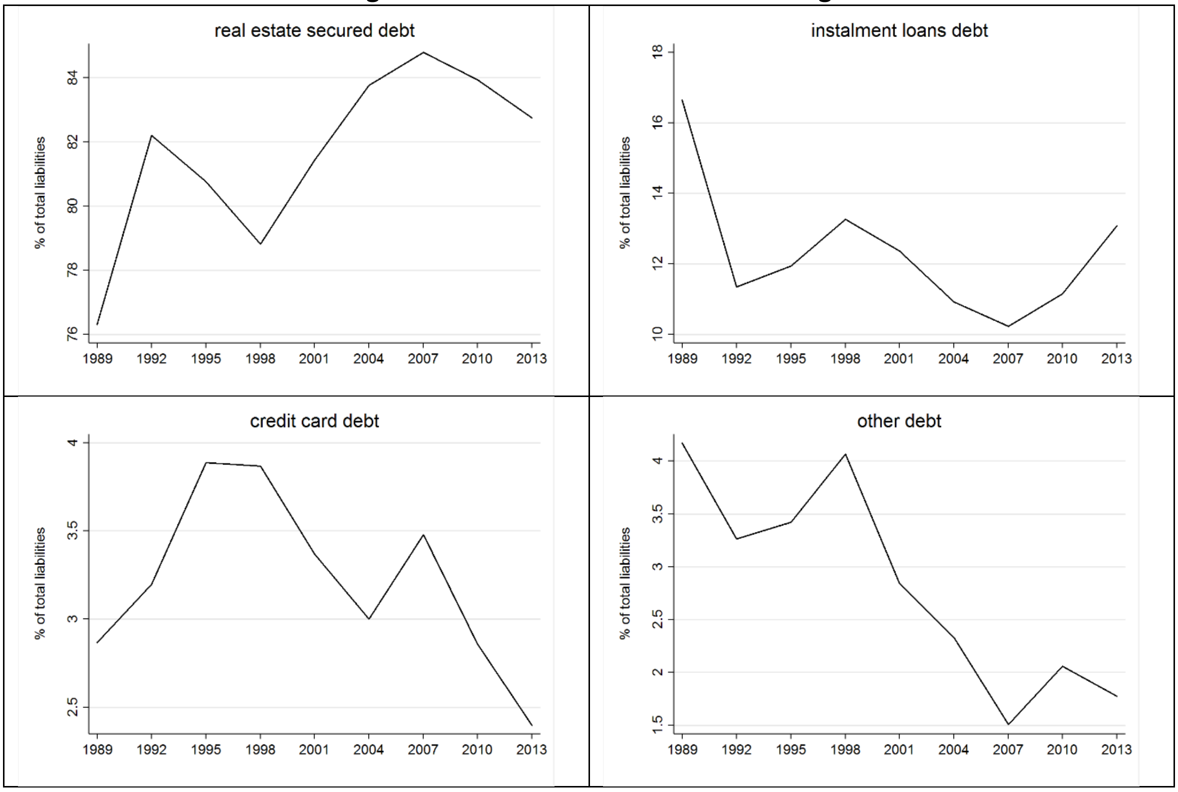
After establishing the increase in US household debt the next step is to split up household liabilities by type. This will prove useful in shedding light on the question of why households took on debt and what did they use the borrowed money for? For example several authors, motivated by the

ECH as well as the ISH, argue that increasing income inequality led to a situation where households borrowed for consumption purposes (Kumhof & Ranciere 2010; Kapeller & Schütz 2014; Ryoo & Kim 2014; Carr & Jayadev 2015; Bertrand & Morse 2013). However the breakdown by type reveals that the vast majority of household debt is mortgage debt and other real estate secured debt. This comes as no surprise since for most households the owner occupied residence is the main asset which is typically purchased by taking out a mortgage. Figure 2 displays the main forms of household borrowing in detail. The upper left hand panel displays the share of real estate secured debt in total outstanding debt. Real estate secured debt includes mortgages and home equity lines of credit. Real estate related forms of borrowing comprise not only more than 80% of total borrowing over the period 1989 to 2013 but its share increased by more than 5 percentage points between 1998 and 2007. This means that over this sub-period mortgages exhibited above average growth compared to other forms of borrowing.

The upper right panel of Figure 2 displays the share of instalment loans, most of which are car and consumer loans as well as education loans. These became less important between 1998 and 2007 and only accounted for roughly 10% of total borrowing in 2007. Since instalment loans are heavily used to finance consumption expenditures the below average growth casts doubts on the expenditure cascades hypothesis as it predicts debt-financed consumption spending as the main reason for household borrowing.

The lower left panel contains the ratio of credit card debt to total outstanding household liabilities. Credit card debt played a minor role for aggregate borrowing and its share in total household liabilities fell from about 4% in 1995 and 1998 to 3% in 2004 and 2.5% in 2013. The share of other forms of borrowing is presented in the lower right panel of Figure 2. Other forms of borrowing include loans against pension accounts and any other liabilities not part of the previous categories. They accounted for less than 4% of total liabilities and their share fell to a low of 1.5% in 2007, indicating that these forms of borrowing grew slower than total borrowing and in particular slower than mortgages. Figure 2 demonstrates that real estate secured borrowing is the elephant in the room, not only because of the actual size of this debt category relative to total outstanding household liabilities but also because of its increasing importance in the years leading up to the financial crisis.

Figure 2: Forms of household borrowing



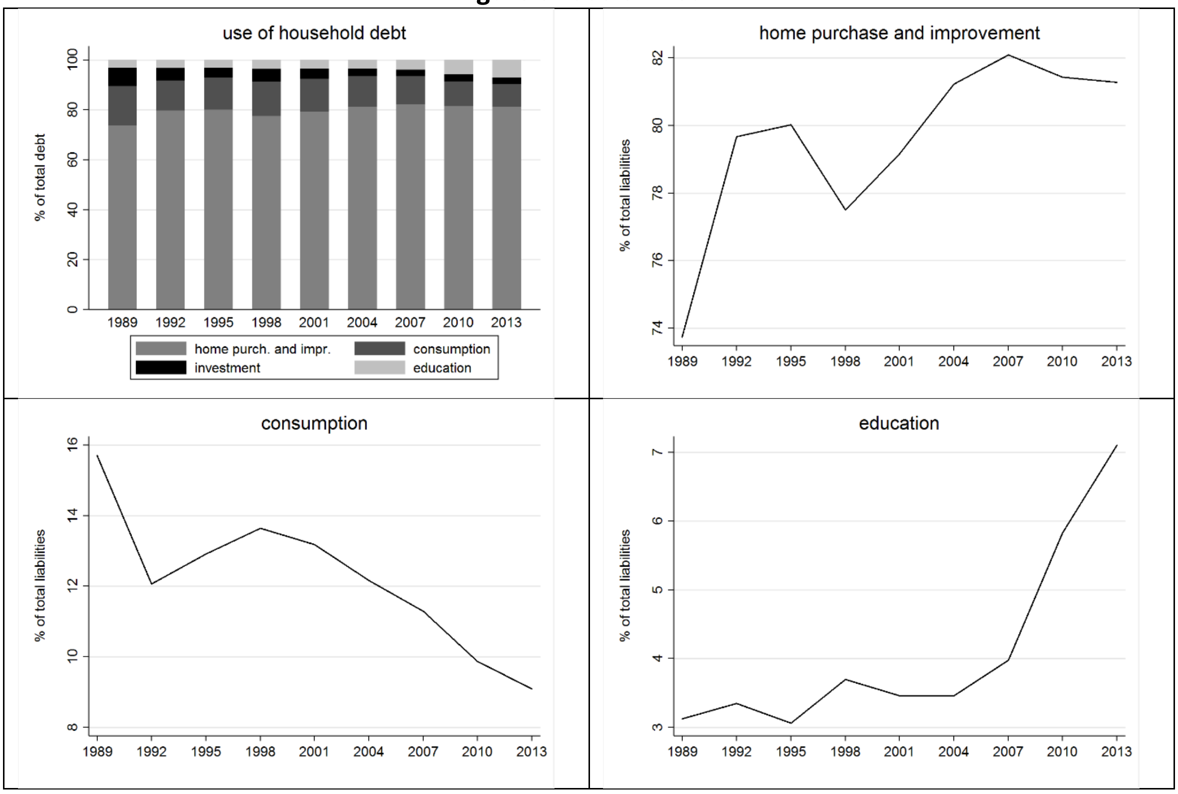
All variables are expressed relative to total outstanding liabilities (see Figure 10 in the Appendix). Source: author's calculations based on SCF waves 1989-2013.

3.4 Uses of Household Credit

Splitting up household debt by types of liabilities provides some information about how the borrowed money was used. However there is not a clean 1:1 relationship between the type and the use of credit. The reason is that a second mortgage for example can still be used to buy a new car. Figure 3 splits up US household debt based on how the money was used. It emerges that households mainly borrowed in order to purchase homes or to improve the current residence. The upper left panel of Figure 3 displays loans for home purchases and home improvements, consumption expenditures, investment purposes, and education expenses in a single graph. It is important to note that purchases of second homes for investment purposes (buy-to-let) are not included in the home purchase and improvement category but in the investment category.

The upper right part of Figure 3 demonstrates that home purchases and improvements motivated 78% of total liabilities in 1998 and 82% in 2007. In comparison, real estate secured debt accounted for about 79% of total liabilities in 1998 and for 85% in 2007. This indicates that a small but increasing proportion of real estate secured debt was used for other purposes than home purchases or improvements. The gap between real estate secured debt and debt taken out for home purchases and improvements amounts to $173 billion in 2001 and $276 billion in 2004. This means real estate secured debt which was not used for home purchases or improvements increased by $103 billion over this period. One obvious way to use such additional borrowing would be for consumption purposes. Greenspan and Kennedy (2008) use a very different method but based on their results, US households used $105 billion (in 2013 Dollars) from refinanced mortgages for consumption between 2002 and 2004[12]. While these numbers are non-trivial, they account only for a small proportion of the total increase in household debt. The lower left panel of Figure 3 might be surprising to some economists because it is in stark contrast with the idea that US households borrowed heavily in order to finance consumption prior to the 2007 crisis. On aggregate between 1998 and 2007 borrowing for consumption purposes declined from almost 14% of total liabilities to about 11%. Of course in absolute terms ($0.94 trillion in 1998 and $1.43 trillion in 2007) and also relative to disposable household income (12.1% in 1998 and 13% in 2007, see Figure 2 in Appendix I) consumption debt did grow over this period, but it grew slower than borrowing for home purchases and home improvements. The lower right panel of Figure 3 reveals a recent development in the United States: the rapid increase in student debt. After 2004 student debt rapidly increased relative to other liabilities. On the one hand this reflects the deleveraging process of the household sector and on the other hand it reflects soaring university tuition fees.

Figure 3: Use of credit



All variables are expressed relative to total outstanding liabilities (see Figure 10 in the Appendix). Source: author's calculations based on SCF waves 1989-2013.

[12] The $105 billion figure from Greenspan and Kennedy (2008) was obtained by using their series 'Free cash generated by refinancings used for private consumption expenditures' which is displayed in line 88 in Table 2. Then these numbers are deflated by the CPI with base 2013 and the years 2002, 2003 and 2004 are summed up.

Altogether the result from the previous section carries over: real estate markets are the main driver of US household borrowing. In particular prior to the financial crisis in 2007, real estate related borrowing was only marginally used for consumption purposes and potentially as a substitute for more expensive forms of borrowing such as unsecured loans and credit cards. These findings are consistent with the housing hypothesis but cast doubt on the ECH as well as the ISH as primary explanations of US household borrowing in the last three decades.

3.5 Income Inequality and Household Borrowing

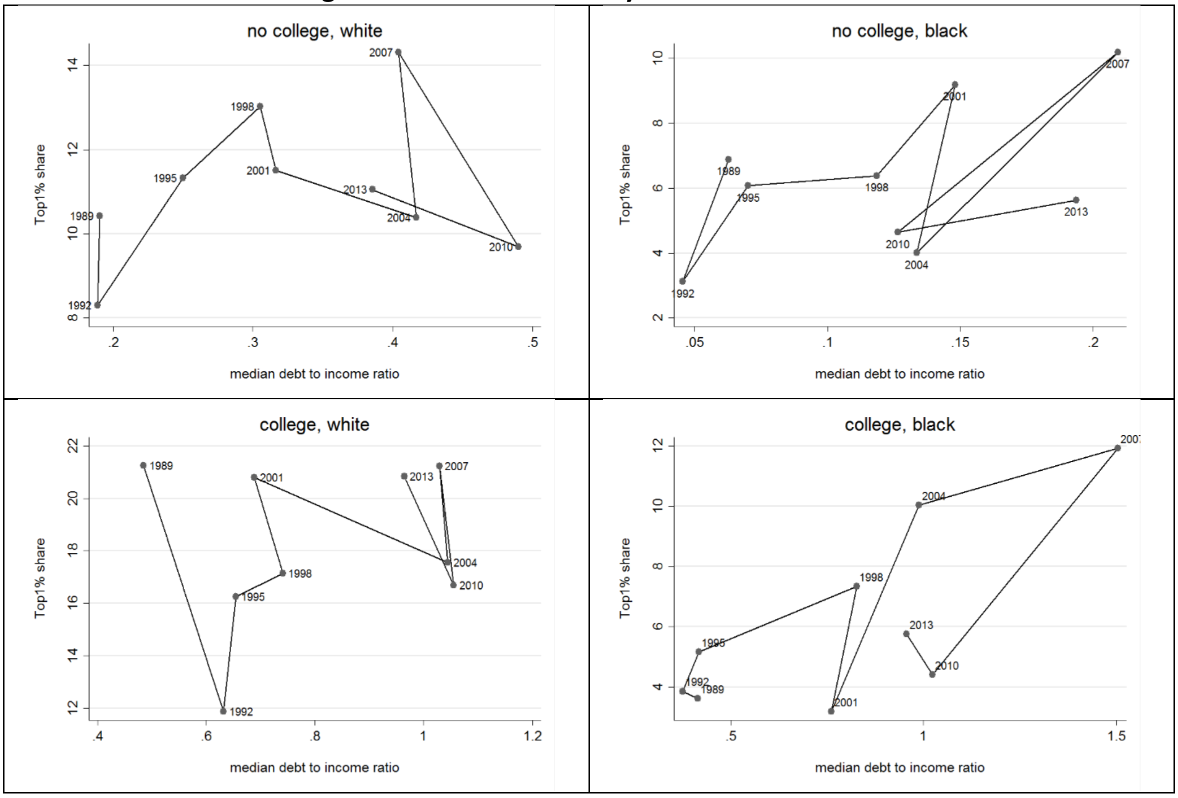
The data presented in the previous two sections is not consistent with the interpretation that US household debt increased due to debt-financed consumption expenditures as a result of growing income polarization. However the question of how and whether the concentration of income at the top of the distribution which occurred over the last 40 years is related to increased household borrowing is a hotly debated issue (Carr & Jayadev 2015; Coibion et al. 2016; Wildauer 2016). This section adds descriptive evidence to that discussion. In order to assess to what extent the data is compatible with the ECH or the ISH it is necessary to clarify within which groups status comparison takes place. The rationale used in this thesis rests on two arguments. First, households compare themselves within their community. Social interaction and signalling of status are strong within the local community. The racial background of the household head is used as a proxy for the area in which a household lives because the US remains highly segregated (Frey 2015; Glaeser & Vigdor 2012). Households are distinguished based on whether the household head is black or white. However the SCF provides four racial categories: white, black, hispanic and other. The reason for not also using households classified as other is that other is a residual category which does not identify a homogeneous group. However reference groups are motivated by the idea that households engage with and compare themselves to similar peers. The reason for dropping the hispanic category is simple: The number of observations within the college-hispanic cells are too low to reliably calculate within group characteristics such as the top 1% income share[13].

The second argument is that households mostly interact with socially similar people. The reason is that people which share a similar background and live experiences are more likely to connect. A good proxy for the 'milieu' in which a household lives, works and spends leisure time is educational achievement. This thesis distinguishes between household heads with a college degree and without a college degree. Together with the racial background that results in four groups: non-college black, non-college white, college black and college white. In order to capture the idea of upward looking status comparison which is an implicit assumption of the ECH, the income share of the richest 1%

[13] For example in 1995 there are only 31 observations in the college-hispanic cell compared to 71 in the college-black cell, 1,602 in college-white, 141 in non-college-hispanic, 285 in no-college-black and 1,839 in no-college-white.

of households within each group is used as a measure of within group income inequality[14]. The idea is that if within group income inequality is high, the households at the top of the income distribution are able to spend heavily on status goods and peer pressure for the bottom 99% increases due to a cascade of status expenditures. Therefore it is expected that rising within group top 1% income shares should coincide with increasing household indebtedness.

Figure 4: Debt to income by race and education



Top income shares are computed separately for each group. Debt to income ratios are computed as the median debt to income ratio among those households below the 99th percentile of the group income distribution. Households with no income are excluded.

It is important to emphasise that the ECH strictly relies on upward looking status comparison. Only if households compare themselves with others higher up in the income distribution will an increase in top income shares lead to a cascade of debt-financed expenditures. In contrast, if households rely on average income as their point of reference, a polarisation in the distribution of income would not lead to an increase in spending and borrowing for those households with above-average incomes. Even though the mean is a commonly used point of reference in the literature (Alpizar et al. 2005; Maurer & Meier 2008; Alvarez-Cuadrado & Vilalta 2012; Alvarez-Cuadrado et al. 2012), it

[14] Within group income inequality refers to the income distribution within a group (e.g. households which identify as black and with a college educated head) in contrast to the aggregate income distribution which refers to the distribution of income within the entire population.

is not consistent with the ECH. For this reason this thesis will not use average income as a point of reference.

Figure 4 plots the median debt to income ratio among those households below the 99th percentile of the within group income distribution against the within group top 1% income share for four groups of US households[15]: non-college black, non-college white, college black and college white. These four groups account for more than 80% of the total outstanding household sector liabilities. The upper left panel contains the trajectory of top income shares and debt to income ratios over time for white households without a college degree. Within this group the income share of the top 1% grew from 8.3% in 1992 to 14.3% in 2007[16]. Over the same period (1992-2007) median debt to income ratios doubled from 0.19 to 0.4 (in absolute terms $1.6 trillion in 1992 and $ 3.7 trillion in 2007, see Figure 4 in Appendix I). The biggest increase in debt to income ratios occurred between 2001 and 2004, a period during which the income share of the top 1% fell from 11.5% to 10.4%. When top income shares jumped 3.9 percentage points between 2004 and 2007 median ratios barely moved. Thus only with a focus on the long run development between 1992 and 2007 one can identify a positive link between top income shares and median income ratios (as well as the absolute amounts of outstanding debt). However for the immediate pre-crisis period (2001-2007) such a link does not exist in the data.

The upper right panel of Figure 4 contains the results for the group of black households without a college degree. Median debt to income ratios (as well as absolute amounts outstanding, see Figure 4 in Appendix I) rose continuously between 1992 and 2007. The important exception was the year 2004 in which median debt to income ratios dropped slightly in the face of sharply falling top income shares. One can think that deleveraging in the face of easing income polarization takes longer than taking on debt as a reaction to rising income inequality which would explain this asymmetric reaction of borrowing to increases and reductions of inequality. In absolute terms the liabilities for this group rose from $147 billion in 1992 to $502 billion in 2007. Between 2004 and 2007 alone debt increased by $148 billion. Based on these results one can argue that there is a pattern of rising income inequality coinciding with increased household borrowing.

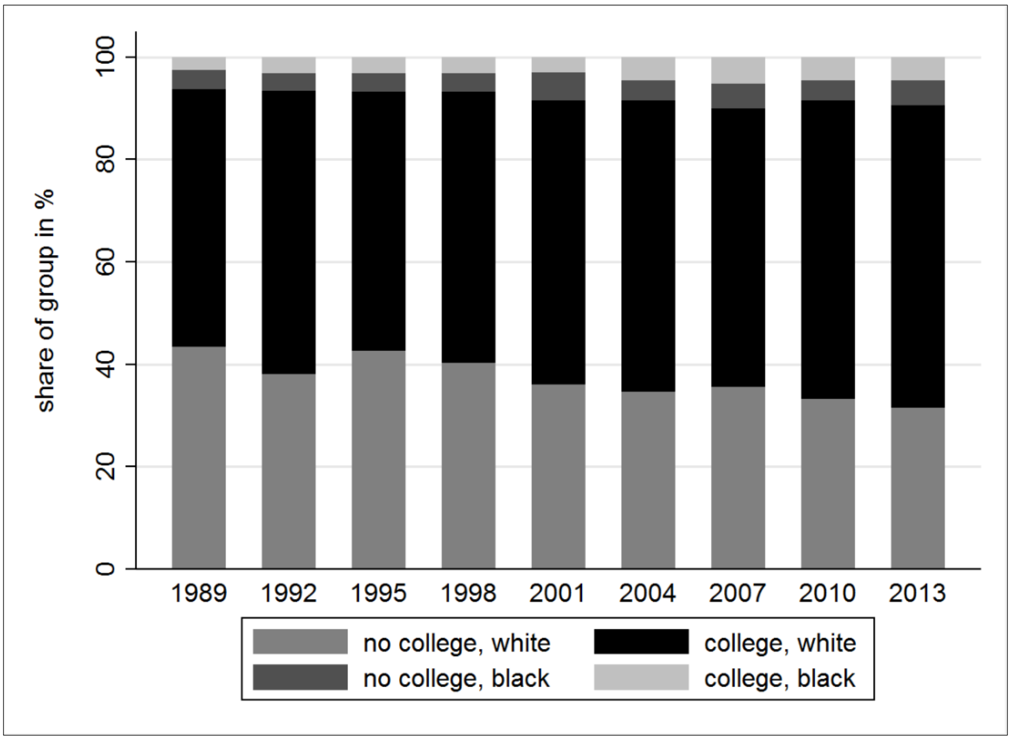
Borrowing habits for the group of white college educated households are strikingly different. Between 1992 and 2001 top income shares increased from 11.9% to 20.8% while the median debt to income ratio barley moved from 0.63 to 0.69. Between 2001 and 2004 debt ratios jumped from 0.69 to 1.04 despite the fact that the top income share fell 5.2 percentage points over this period.

---

[15] Qualitatively equivalent results emerge if the top 5% income share is used. See Figure 3 in Appendix I.
[16] Within group top income shares provide a normalized measure of within group inequality. However they do not provide any information about between group differences. Table 1 in Appendix I contains the average income for the top 1% for each group which can be used as an indicator for between group inequalities.

Also in absolute numbers the biggest increase in outstanding liabilities occurred in that period (from $3.7 trillion in 2001 to $5.1 trillion in 2004). Similar to the group of white non college educated households there is a positive longer term relationship between top income shares and household borrowing (measured by median debt to income ratios as well as absolute amounts) but for the period of 2001-2007 such a relationship is clearly not present in the data. Put differently, the data is not consistent with the EC as an explanation of borrowing behaviour of white college educated households in the pre-crisis period.

Figure 5: Debt by race and education, relative sizes



The Figure presents the amount of total outstanding liabilities within each group as % of the total outstanding liabilities of all four groups taken together.

The fourth panel of Figure 4 contains the results for black college-educated households. Similar to the group of non-college educated households over the 1992-2007 period there is a clear pattern of rising top income shares and rising median debt to income ratios. A similar pattern arises from the absolute amounts borrowed. Between 1998 and 2001 top income shares fell and so did the median debt to income ratio. Even though the reduction seems small given the size of the drop of top income shares, the direct relationship still holds and the asymmetry between periods of increasing and decreasing income polarization might be explained by the fact that it takes households longer to deleverage than to leverage up. In absolute terms the outstanding liabilities of this group increased from $133 billion in 1992 to $549 in 2007 (see Figure 11 in the Appendix) while the income share of the top 1% grew from 3.9% to 11.9%. The pronounced jumps in debt ratios which coincide with dramatic increases in income inequality support the expenditure cascades hypothesis as an explanation of this group's borrowing behaviour.

From the data presented in Figure 4 an important conclusion follows: The expenditure cascades hypothesis is not a convincing explanation for the pre-crisis increase in household debt. This conclusion holds even though the data for black households (college and non-college educated) is consistent with the expenditure cascades hypothesis. However the limited size of these two groups does not allow to generalize this conclusion to the US as a whole. The relative size of each group in terms of outstanding debt is presented in Figure 5 above.

3.6 Borrowing as a Reaction to Stagnant Income Growth

Up until now the role of the distribution of income for household borrowing was only considered together with the ECH. However one can also argue that if incomes of low and middle income households stagnate, these households will try to avoid cutting down consumption expenditures by taking on debt in order to maintain their status of living. This argument seems particularly plausible when key expenditures such as housing (or university fees, for some households), grow faster than real incomes.

In order to obtain an idea to what extent stagnant or even negative income growth is an important factor for household borrowing, Figure 9 displays the share of total debt owed by those households which report that their income is lower compared to 'normal' periods in the past. The proportion of total liabilities owed by households with lower than normal income increased from 12% ($0.8 trillion) in 1998 to 18% ($2 trillion) in 2004. Between 2004 and 2007 this proportion fell back from 18% ($2 trillion) to about 15% ($1.9 trillion) and peaked in 2010 at more than 25% ($3.1 trillion) of total liabilities. Some of the volatility of this measure during the 2004-2010 period is explained by business cycle fluctuations. It is also important to note that the income measure the SCF uses is related to last year's income[17]. The decline in the debt proportion of households with lower than normal incomes between 2004 and 2007 despite stable absolute amounts is interpreted as a sign that between 2004 and 2007 the US economy grew quite strongly and thus less people reported that their incomes are lower than usual. At the same time other factors, mainly rapidly growing house prices, encouraged borrowing especially up to the peak of the housing bubble in 2006 (and the 2007 income data actually refers to income in 2006). The jump between 2007 and 2010 then reflects the impact of the recession which hit the US in 2009. Due to the recession more people faced lower than normal incomes and therefore the share of liabilities they were holding jumped.

[17] However assets and liabilities are reported based on current market values. This is an unavoidable inconsistency in the data.

Figure 6: Outstanding liabilities of households hit by negative income shocks

Source: Author's calculations based on SCF waves 1989-2013.

Total household debt increased by $5.8 trillion between 1998 and 2007. Over the same period, debt holdings by households with lower than normal incomes increased by $1.1 trillion. While this is a significant part of the total increase, it cannot explain the majority of the increase. This means that rising income inequality due to stagnant low and middle income growth does not explain the US experience prior to 2007 as predicted by the ISH. This does not mean that such a mechanism did not exist or was not important for some communities but it cannot be seen as the primary explanation of US household borrowing.

The data presented in this and the previous section does not support the notion that the ECH or the ISH can be regarded as a primary explanations for rising aggregate debt levels. Other theories need to be taken into account also because the analysis so far is simply based on unconditional information. This means that even if there were a direct relationship between debt and inequality this might just be a coincidence because the actually relevant factor is omitted[18]. Two alternative explanations are especially important. The first factor which might be missing is the housing market and the second one is credit supply conditions. The next two sections will address these two issues.

3.7 Real Estate Wealth and Household Debt Levels

In an environment of rising house prices first time buyers will have to take out larger mortgages if they do not want to postpone their purchase. Also homeowners will experience capital gains which might lead them to take out additional debt in order to 'extract' some of their home equity and positive home equity eases borrowing constraints. In addition rising property prices will trigger

[18] Which is the motivation for applying regression analysis in order to take several factors into account simultaneously (Wildauer 2016).

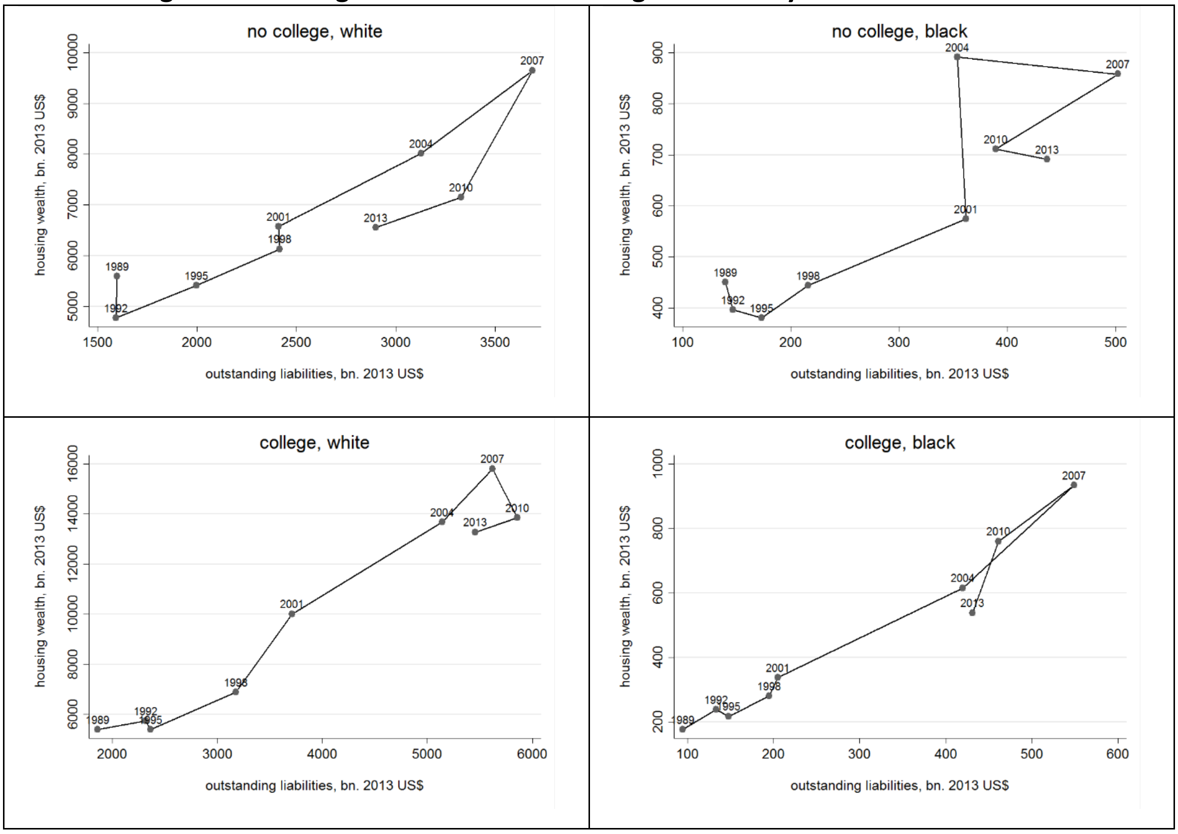
debt-financed property purchases if further price increases and thus potentially substantial capital gains are expected. It is these mechanisms which might be fundamental for understanding household borrowing over the 1989-2007 period. In order to make the results of this section comparable to the results in section 3.5 the connection between housing wealth and household borrowing is studied for each of the four groups. Figure 6 displays the co-movement of housing wealth and total outstanding liabilities for each of them. The measure of housing wealth used in Figure 7 includes the value of primary residences and any other residential real estate held by each group.

For the group of white non college educated households there is a clear direct relationship between the two for most of the sample period. The 2001-2004 period is of special interest because it was this period during which a substantial increase in median debt to income ratios occurred despite falling top income shares. The increase of housing wealth by $1.4 trillion over that period coincided with an increase in outstanding liabilities of $717 billion. Between 2004 and 2007 the corresponding amounts are a $1.6 trillion increase in housing wealth and an increase in debt of $560 billion. These numbers strongly point to the real estate market as the driving force of debt accumulation for white non-college educated households, especially in the 2001-2007 period.

For black non college educated households the direct relationship between housing wealth and outstanding liabilities is not as clear as for the previous group. The periods 2001-2004 and 2004-2007 are surprising in the sense that the value of the groups' housing wealth rose considerably but only afterwards outstanding liabilities increased. The fact that data is only available at 3 year intervals, complicates the analysis because much information is missing. The relatively low value of housing wealth in 2007 most likely reflects the fact that house prices in the US began to fall already in 2006 but deleveraging would not have begun on a large scale before 2007. Nevertheless also for this group the strong connection between real estate wealth and household debt is clearly visible in the data.

The group of white college educated households which is presented in the lower left panel of Figure 7 displays a clear pattern of liabilities rising with real estate wealth. Between 2001 and 2004 the outstanding liabilities of this group increased by $1.4 trillion while the value of the group's residential real estate rose by $3.7 trillion. It is hard to believe that the period with the largest increase in real estate wealth in the sample period coincides only by chance with the largest increase in outstanding liabilities. The lower right panel presents results for the group of black college educated households. Also for this group the biggest increase of outstanding debt which occurred between 2001 and 2007 coincided with a spectacular expansion of housing wealth.

Figure 7: Housing wealth and outstanding liabilities by race and education

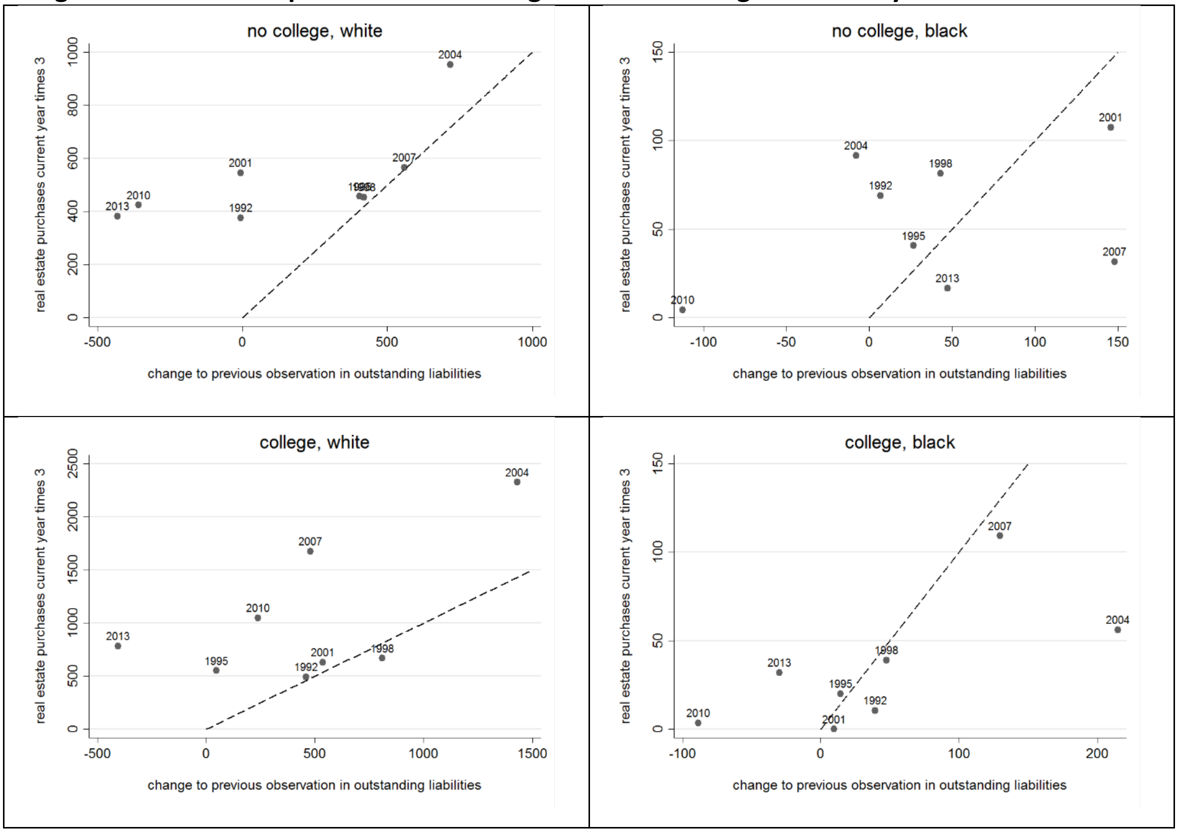


Liabilities are computed as total outstanding liabilities for the bottom 99% of each group. Housing wealth is computed as the value of all residential real estate held by the bottom 99% of each group. All monetary variables are expressed in constant 2013 Dollar values. (See Figure 5 in Appendix I for a graph with median debt to income ratios instead of absolute values of outstanding liabilities).

The fact that increases in debt coincided with increases of housing wealth does not reveal the mechanism of borrowing. In particular it remains unclear whether liabilities increased in order to purchase homes or whether homeowners extracted equity as a result of the capital gains they experienced. Figure 8 sheds some light on the role and size of real estate purchases. It plots the change in outstanding debt against the value of residential real estate purchases which occurred in the year of the survey. Because data is only available in three year intervals and because the changes in debt reflect 3 year changes the value of real estate purchases was multiplied by three. In addition each panel contains a 45 degree line which marks all points where the value of purchases equals the change in liabilities. So if the value of purchases measured for example in 2004 is equal to the value of purchases in 2003 and 2002 and if households only relied on debt to finance these purchases then the 2004 data point would lie on the dashed line. The dashed line is not supposed to indicate a realistic scenario of household behaviour but is supposed to make the interpretation of the graphs easier. Due to the fact that most households will not rely entirely on debt when purchasing residential real estate one expects that the data points will lie above the dashed line. Accordingly if the value of annual purchases in the year of the survey (e.g. 2004) are much higher than in the previous two years (e.g. 2003 and 2002), then taking three times the value from the current survey year (e.g. 2004) will overstate the value of purchases for that three year period and the data points will lie below the dashed line. Similarly if the purchase values of the current year underestimate the purchases in the previous years then the data point is expected to lie above the line.

The pattern which emerges from Figure 8 is that large increases in household debt coincided with high volumes of real estate purchases. An important exception are the post crisis years of 2010 and 2013 which are characterised by relatively low purchase volumes and declining debt volumes. For both groups of white households the data points for the crucial years of 2004 and 2007 are in the upper right corner of the diagram which indicates that the largest increases in household liabilities were associated with very high volumes of purchases. Measurement problems due to the three year intervals are an issue but the general conclusion that debt increased mainly due to real estate purchases holds. Together with the evidence from Figure 6 it becomes clear that in order to explain the aggregate increase in household liabilities, rising house prices and real estate purchases emerge as the most important factor, especially for the pre-crisis period of 2001-2007.

Figure 8: Real estate purchases and changes in outstanding liabilities by race and education



Changes in liabilities are computed as the change relative to the previous observation and thus are three year changes. Real estate purchases correspond to purchases in the year of the survey and are multiplied by 3. All monetary variables are expressed in constant 2013 Dollar values. The dotted line represents all points where real estate purchases equal the change in liabilities.

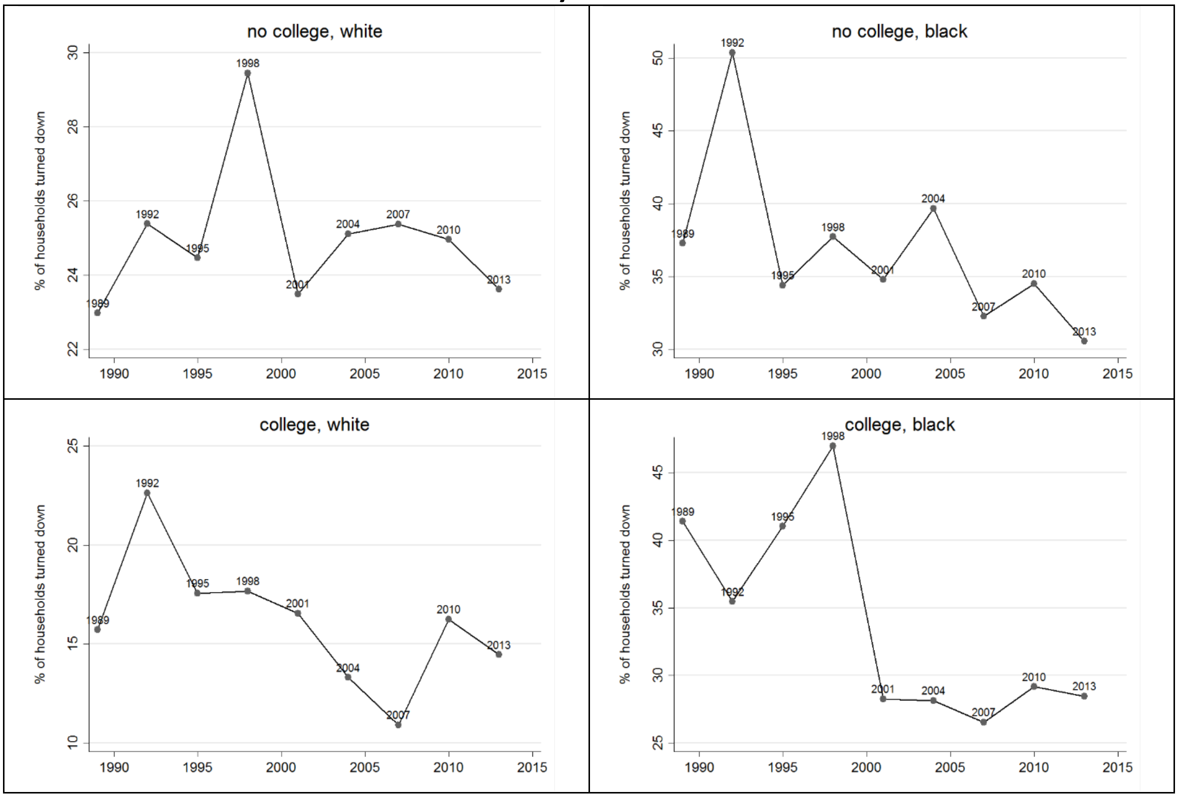
3.8 The Financial Sector's Willingness to Lend

The second important factor which was missing from the analysis so far is the financial sector and potential shifts in credit supply conditions. Changing attitudes in the financial industry and in particular an increasing willingness to lend can be an important factor in explaining the expansion of household debt up until 2007. The emerging practice of securitization and re-securitization of mortgages allowed financial institutions to quickly offload risks from their balance sheets which in turn allowed them to become much more willing to lend to low quality debtors. Since the SCF is a household survey it does not contain information about banking practices but it does contain information on whether a household was rejected when it applied for credit in the past. Several authors document the decline in banks' lending standards (Kiff & Mills 2007; Dell'Ariccia et al. 2012) and Purnanandam (2011) links this decline in lending standards to the rise of the originate-to-distribute model of banking. Purnanandam's argument is that financial institutions adopted a business model which by securitizing mortgages they have given out and by selling the resulting assets, allowed them to pass on a large part of the credit risk associated with traditional mortgage banking to financial markets. Because they could (seemingly) lay of part of the credit risk associated with (subprime) mortgages they started to lower their lending standards and thus granted borrowers with very low credit ratings access to the mortgage market. Against this background declining credit application rates can be interpreted as an indicator of changing bank behaviour, representing a general relaxation of credit standards.

While observing that fewer households have their credit applications turned down could also result from a general improvement in the quality of credit applications, it is not clear why the overall quality of applicants should improve in a given year. The reason is because even in an economic boom period in which households' disposable income grows strongly, new applicants with low credit scores will try to borrow, encouraged by the general positive economic circumstances. Thus, the ratio of turned down borrowers is interpreted as an indicator of credit supply conditions, assuming that general shifts in the quality of applicants do not drive this ratio.

Figure 9 presents the number of rejected credit applicants relative to the total number of debtors for each group over the sample period. The fact that the number of rejected households are presented relative to households holding any form of debt takes into account that as more people take on debt also the number of turned down households would increase even if credit standards remain unchanged. So using the (increasing) number of actual debtors as the scaling factor takes this effect into account. Figure 9 shows that for three groups the relative number of households whose credit applications were rejected fell over time. For the group of white college educated households the ratio declined from 22% in 1992 to 11% in 2007. For the other groups the trends are more volatile. For non college educated black households the proportion of households turned down decreased from 37% in 1998 to about 32% in 2007. Despite pronounced volatility there is a downward trend over the entire sample period. A similar observation holds for the group of college educated black households. Starting out with a rejection rate of about 42% in 1989 by 2007 this rate went down to roughly 26%. The only group for which the proportion of rejected households does not fall considerably over the sample period are non-college educated white households. Between 2001 and 2007 the rejection rate even increased from almost 24% to almost 26%. Overall the data supports the argument that shifts in the willingness to lend by the financial sector were an important factor which allowed households to leverage up over the sample period.

Figure 9: Proportion of households whose credit application was turned down within last 5 years



The proportion of turned down households is computed as the share of those households within each group who were turned down when applying for credit over the last 5 years relative to those households holding any debt.

3.9 Summary of Chapter 3

The results of this chapter can be summarised as follows: First, US households increased their liabilities substantially. Between 1989 and 2007 total outstanding liabilities of the household sector rose from \$4.4 trillion to \$12.7 trillion in real terms. This increase was the result of two effects: households took on more debt and more households took on debt. The first effect is reflected by the increase of the median debt to income ratio (including only indebted households) from 51% in 1989 to 111% in 2007. The second effect is reflected by the increasing proportion of households with any form of liability (72.3% in 1989 and 77% in 2007).

Second, the data presented strongly points to the dominant role of the housing market as a driving factor of household debt accumulation. More than 80% of outstanding household liabilities are mortgages, most of which were taken out to purchase or improve property and not to spend otherwise. Especially in the 2001-2007 period household debt and housing wealth expanded in parallel. This is interpreted as an interplay of the following three developments: first, aspiring homeowners being forced to buy into a higher priced market; second, homeowners taking on debt to extract equity; and third, risk-seeking households making debt financed purchases to speculate on future capital gains. The chapter does not attempt to distinguish between these channels. Nevertheless these results can be interpreted as being strongly consistent with the housing hypothesis.

Third, the nexus between intra group income inequality and household indebtedness is shaky. The data supports such a relationship for black households but not for white households. Especially in the pre-crisis period between 2001 and 2007 top income shares and household indebtedness did not move in parallel. This is interpreted as the data not supporting the ECH as an explanation of increasing household indebtedness and thus in contrast with several recent theoretical (Belabed et al. 2013; Kapeller & Schütz 2014; Kumhof & Ranciere 2010; Ryoo & Kim 2014) as well as empirical (Bertrand & Morse 2013; Carr & Jayadev 2015) studies using survey data.

Fourth, the notion that borrowing was concentrated among households which suffered income declines relative to the past and used debt in order to maintain living standards (Barba & Pivetti 2009) is not confirmed by the data. Borrowing by households with less than normal incomes was non-trivial but accounts only for a fifth of the total increase in household debt prior to the crisis. Therefore, the ISH by no means can serve as the main explanation for the increase in household liabilities up to 2007.

Fifth, shifts in credit supply conditions seem to have some explanatory power especially for the group of white college educated households. For this group, the share of households which were rejected when applying for credit dropped from the peak of 22.6% in 1992 to 10.9% in 2007.

So overall the analysis carried out in this chapter indicates that the housing hypothesis is more in line with the data on US household debt levels compared to the expenditure cascades and the income stagnation hypothesis. It seems that the HH is able to explain a large part of the increased household sector indebtedness while the ECH and the ISH can be applied to subsamples only (ECH to black households) or only explain a small proportion of the increase in household debt (ISH).

Nevertheless, it is important to stress that the analysis presented in this chapter is not always able to make a clean distinction between the individual explanations. For example if credit supply conditions ease for specific groups which are clustered in certain areas then easing credit conditions will allow house prices in that area to grow. However despite these complications the results presented yield a clear and strong conclusion: Housing is the key factor in household debt dynamics. Any attempt to explain or model the leverage of the US household sector without taking the housing market into account is missing the point. This is interpreted as strong support for the housing hypothesis.

This is not to say that the ECH or the ISH should be entirely ignored. A key feature of financial crises is that a collapse of one market segment can have big implications if a panic develops and (institutional) investors seek to repair their balance sheets by selling off other assets. Such fire sales lead to plummeting financial asset prices and cause problems to spread quickly across the financial sector. The IMF's former chief economist emphasizes that subprime mortgage defaults of about $250 billion in 2007 where the trigger for the financial crisis (Blanchard 2009). This is important because the absolute amounts of borrowing by black households which coincide with rising top income shares and the amounts borrowed by households facing less than normal income levels are non-trivial and are large enough to have macroeconomic effects, especially when it comes to triggering a credit crisis. Total liabilities of black households amounted to $1.05 trillion out of $12.7 trillion and thus to 8.3% of total liabilities in 2007. Borrowing by households with less than normal incomes amounted to $1.9 trillion in 2007.

4 Determinants of US Household Debt: New Evidence from the SCF

This chapter specifically focuses on the relative explanatory power of the ECH and the HH. With spectacular increases in US property prices and record levels of income inequality prior to the financial crisis, both hypotheses seem to be logically consistent with the rise in US household debt. This chapter is the first attempt in the literature to assess the explanatory power of both explanations in a unified framework. The most relevant studies which rely on survey data and are interested in the determinants of household debt, focus only on one of these two hypotheses.

Carr and Jayadev (2015) for example are primarily interested in the impact of the households' relative position within the income distribution on household borrowing. Since their model controls for income separately they interpret a positive impact of the relative position as consistent with the ECH. They use data from the Panel Study of Income Dynamics (PSID) from 1999 to 2009 and do not control for individual real estate wealth or local property prices but only for homeownership. Coibion et al. (2016) use data from the Consumer Credit Panel and the SCF and they find that debt accumulation relative to income was lower for low income households living in a high income inequality area in 2001 compared to low inequality areas. Equivalently, high income households in high inequality areas borrowed more compared to high income households in low inequality areas. Coibion et al. (2016) do not control for real estate wealth separately but argue that the panel structure allows them to control for unobserved heterogeneity. Also closely related to this chapter is the work of Mian and Sufi (2011). They analyse the variation across metropolitan statistical areas in household debt growth for homeowners between 2002 and 2006. Mian and Sufi (2011) find large effects of house prices on borrowing especially for homeowners with low credit scores and high propensities to borrow on credit cards which they interpret as evidence of credit constraints and self-control problems, respectively. However they do not take into account any potential distributional effects as would be implied by the ECH. Overall this brief discussion of the related literature demonstrates that a comprehensive study is missing. In addition Carr and Jayadev (2015) as well as Mian and Sufi (2011) rely entirely on data which is unlikely to adequately capture households at the top of the income distribution and thus miss a part of the US population which is crucial when it comes to analysing consequences of distributional shifts. In addition to analysing the effects of real estate wealth and the distribution of income in the same framework, this chapter uses SCF data and benefits from this data's unmatched coverage of the tail of the income and wealth distributions. In order to make SCF data useful for an econometric analysis the chapter develops a method for deriving annual changes in household liabilities, despite the dataset's lack of a panel structure.

The contribution to the existing literature is threefold: First, deriving annual household borrowing from the SCF allows the author to investigate the accumulation of household debt based on the best available survey data on household finances. The unmatched advantage of the SCF compared to other US household surveys is that it relies on tax information in order to adequately capture the top tail of the income and wealth distribution and thus is the only survey which deals in a convincing way with non-observation and differential non-response problems. In addition, among all US household surveys, the SCF provides the most detailed breakdown of the household balance sheet. Detailed balance sheet information is important because current asset holdings and currently outstanding liabilities are crucial determinants of current household borrowing. Second, the chapter brings together two powerful explanations of US household borrowing by assessing the impact of income inequality and rising property prices in a unified model. Doing so not only avoids potentially serious omitted variable problems but also allows a direct comparison of the effect size of both hypotheses. Third, this is the first attempt in the literature to separately analyse the groups of borrowing and non-borrowing households. Fundamentally different results for the two groups demonstrate that borrowing and debt repayment decisions are not symmetrical. Separating them in the analysis therefore yields a better description of household behaviour which is reflected in an improved model fit.

Based on the newly developed measure of household borrowing the chapter finds strong evidence supporting the HH and support for the ECH conditional on homeownership. This is interpreted as households having been able to engage in status driven debt-financed expenditures only if they had access to collateral. Households borrowed $963 billion[19] more in 2004 compared to 1995 due to the increase in residential real estate prices and the rise of top incomes. In addition, the results also indicate that rising real estate prices strongly contributed to household indebtedness via home purchases: households borrowed $584 billion more in 2004 compared to 1995 due to rising costs of purchases. These findings indicate that even if income inequality and relative incomes play a role for household borrowing, because of credit constraints, the prices of real estate assets and thus the housing market is the dominant factor when it comes to explaining household borrowing outcomes.

The rest of the chapter is organized as follows: section 4.1 develops the method used to compute the change in household liabilities from the cross-sectional dataset, section 4.2 presents the econometric model as well as the definitions of reference income, section 4.3 presents the estimation results and section 4.4 reports aggregate level effect size computations. A final section summarises the results.

4.1 The Problem of the Missing Time Dimension in the SCF

The analysis of this chapter is based on SCF data. The raw dataset used covers the years 1995 to 2007 in three-year intervals and consists of 21,982 original observations. Since the SCF is a multiply imputed dataset with five imputations (see discussion in section 3.1.3), the total number of

[19] All monetary values in this chapter are expressed in 2013 prices.

datapoints in the raw dataset is 109,910. 810 datapoints with zero income are dropped as well as 2,460 datapoints which exhibit a negative debt level in the previous period. The SCF does not provide information on how much additional money was borrowed on the first mortgage on the primary residence in case of equity withdrawals in the surveys prior to 2004. Thus in 1995, 1998 and 2001 a total of 748 datapoints are excluded, falling into that category[20]. In addition 294 datapoints which exhibit declines in outstanding liabilities of more than 250% of disposable income as well as observations of increases in liabilities of more than 500% of disposable income are excluded. This leaves a final sample of 102,958 datapoints corresponding to roughly 20,592 observations per implicate. Sample sizes reported in regression tables refer to the number of observations per implicate.

A key limitation when using the SCF in econometric analysis is the missing time dimension of the dataset due to its design as a repeated cross section. Unlike in a panel one cannot follow the same household through time. However the detailed information the survey collects about the credit history of each participant allows to infer by how much the outstanding liabilities changed within the year of the survey. Thus one can infer whether an individual household took on debt in the year of the survey, paid back already existing liabilities or pursued neither of these activities. Creating such a measure is key for answering the question whether households which are exposed to rich peers take on debt to keep up in spending with these peers. Without knowing whether and by how much households take on debt in each year one would have to rely on the total amount of outstanding liabilities. The stock of debt however is strongly influenced by past decisions, most importantly real estate purchases, and is therefore not a good measure of whether households which are currently exposed to richer peers take on debt. This section presents the steps involved in creating a measure of the within year change in household liabilities.

In order to understand how the change of an individual household's debt level is constructed one has to keep in mind that the SCF covers 10 different debt categories. Participating households are asked about their outstanding liabilities with respect to mortgages (primary residence as well as other properties), lines of credit, credit on land contracts, consumer loans, credit cards, car and vehicle loans, education loans, loans against pension plans and other loans. Total outstanding liabilities of a household in the sample (D)[21] consists of up to three outstanding mortgages on the primary residence (D^{M1} to D^{M3})[22], three credit lines (D^{CL1} to D^{CL3})[23], outstanding land contracts

[20] Households which extract home equity but do not have a mortgage prior to the equity extraction are still part of the sample.
[21] In the 2013 summary dataset this variable is called "debt".
[22] In the 2013 codebook (https://www.federalreserve.gov/econres/files/codebk2013.txt) questions X805, X905 and X1005 ask about the currently outstanding amounts.
[23] In the 2013 codebook questions X1108, X1119 and X1130 ask about the currently outstanding amounts.

(D^{LC})[24] which are loans on property the household no longer owns, up to three mortgages on other properties (D^{OM1} to D^{OM3})[25], up to six consumer loans (D^{CL1} to D^{CL6})[26], up to five outstanding credit cards (D^{CC1} to D^{CC5})[27], up to four car loans (D^{CAR1} to D^{CAR4})[28], up to six education loans (D^{EDU1} to D^{EDU6})[29], loans against pension plans (D^{P})[30] and other loans (loans to purchase property (D^{OLP})[31] and loans for home improvement (D^{HI})[32]). So for every household i, the total amount of outstanding liabilities in year t is the sum of all outstanding liabilities on all these items:

$$D_{i,t} = D_{i,t}^{M1} + D_{i,t}^{M2} + D_{i,t}^{M3} + D_{i,t}^{CL1} + D_{i,t}^{CL2} + D_{i,t}^{CL3} + \cdots + D_{i,t}^{OLP} + D_{i,t}^{HIP} \quad (1)$$

While the outstanding amount for each of these categories is collected by the survey, the SCF does not provide information on how much these categories changed within the year of the interview. Thus, the SCF does not provide information on the amount of borrowing at the household level ($\Delta D_{i,t}$). However based on the detailed information the survey collects in all these categories, the author can construct a measure of change for individual categories and by summing up the changes across categories an estimate for borrowing at the household level is obtained. For four categories, the survey does not collect enough information to make such an inference: credit card debt (D^{CC1} to D^{CC5}), credit lines (D^{CL1} to D^{CL3}), loans on land contracts (D^{LC}) and loans against pension plans (D^{P}) and thus changes in these categories are not included in the measure of household level borrowing derived in this section.

In order to understand how changes for each category of liabilities were derived the process will be discussed in detail for the first mortgage on the primary residence. The outstanding amount for household i in year t will be denoted $D_{i,t}^{M1}$ where $i = 1,2, \dots, N$ and $t = 1995,1998, \dots ,2007$ indicates the year in which the household was interviewed. The amount initially borrowed on this first mortgage is denoted $B_{i,j}^{M1}$ where $j = 1944,1945, \dots ,2007$ indicates the year in which this mortgage was taken out. Understanding by how much the outstanding amount on the first mortgage changed in the year of the interview involves two steps. The first step distinguishes whether the mortgage was taken out in the year of the interview ($t = j$) or prior to the year of the interview ($t > j$) and how the money was used (use). The use of the money is summarized in four

[24] In the 2013 codebook questions X1318 and X1337 ask about the currently outstanding amounts.
[25] In the 2013 codebook questions X1715, X1815 and X1915 ask about the currently outstanding amount.
[26] In the 2013 codebook questions X2723, X2740, X2823, X2840, X2923 and X2940 ask about the currently outstanding amount.
[27] In the 2013 codebook questions X413, X421, X424, X427and X430 ask about the currently outstanding amount.
[28] In the 2013 codebook questions X2218, X2318, X2418 and X7169 ask about the currently outstanding amount.
[29] In the 2013 codebook questions X7824, X7847, X7870, X7924, X7947 and X7970 ask about the currently outstanding amount.
[30] In the 2013 codebook question X11027 asks about the currently outstanding amount.
[31] In the 2013 codebook question X1044 asks about the currently outstanding amount.
[32] In the 2013 codebook question X1215 asks about the currently outstanding amount.

categories[33]: refinancing of an earlier mortgage and thus not engage in additional borrowing ($use = ref$), borrowing additional money and thus extracting equity ($use = extr$), extracting equity and refinancing an earlier mortgage ($use = extr + ref$) and taking out a completely new mortgage with no prior mortgage ($use = no\ prior\ mortgage$).

In cases where the mortgage was taken out in the year of the interview ($t = j$: cases 1.1 to 1.4) the information on the use of the mortgage becomes crucial. If the mortgage was used to refinance an earlier credit ($t = j$ and $use = ref$: case 1.1), the change in the outstanding balance of that mortgage can be derived as the difference between the amount currently outstanding ($D_{i,t}^{M1}$) and the amount initially borrowed ($B_{i,t}^{M1}$): $\Delta D_{i,t}^{M1} = D_{i,t}^{M1} - B_{i,t}^{M1}$. The rationale for this derivation is that the amount initially borrowed does not constitute a new liability but replaced an already existing one. Thus, because the mortgage was taken out in the year of the interview ($t = j$) only the difference between the refinanced amount and the currently outstanding amount represents an actual change in outstanding liabilities. It is important to note that depending on whether the amount initially borrowed is smaller, bigger or equal compared to the amount currently outstanding, the resulting change will be positive, negative or zero. The case of $B_{i,t}^{M1} < D_{i,t}^{M1}$ is interpreted as household i being behind on payments and accumulating overdue interest which is added to the total amount due and thus the currently outstanding amount exceeds the amount initially borrowed.

In contrast if the mortgage was used to extract equity from the property ($t = j$ and $use = extr$: case 1.2), the change in the amount outstanding is defined as the amount extracted (ex_{it}): $\Delta D_{i,t}^{M1} = ex_{it}$ which represents newly accumulated debt.

If the mortgage was taken out to extract equity and to refinance an earlier loan ($t = j$ and $use = extr + ref$: case 1.3) the change is defined as the amount extracted plus the difference between the amount currently outstanding and the amount initially borrowed: $\Delta D_{i,t}^{M1} = ex_{it} - \left(B_{i,t}^{M1} - D_{i,t}^{M1}\right)$. The difference between the initial and the current amount is subtracted because this difference represents the extent to which (new as well as already existing) debt was paid down.

If the household had no prior loan or mortgage ($t = j$ and $use = no\ prior\ mortgage$: case 1.4) the change in debt is simply defined as the amount currently outstanding because the amount currently outstanding represents debt accumulated in the current period: $\Delta D_{i,t}^{M1} = D_{i,t}^{M1}$. The reasoning is that in this case $D_{i,t-1}^{M1} = 0$ and thus the change in debt equals the amount currently outstanding.

[33] In the 2013 codebook question X7137 asks about the use of this mortgage.

The second step involves distinguishing households which did not take out their mortgage in the year of the interview ($t > j$: cases 2.1 to 2.3). Thus, for these households the task is to infer the amount of principal which has been repaid over the year in which the interview was taken. This inference is made based on the answer to the question whether the initial amount borrowed ($B_{i,j}^{M1}$) is bigger, equal or smaller than the amount currently outstanding ($D_{i,t}^{M1}$). In case the amount initially borrowed is bigger than the amount currently outstanding ($t > j$ and $B_{i,j}^{M1} > D_{i,t}^{M1}$: case 2.1) it is assumed that these households are paying back their mortgage on schedule. Thus the change in debt equals the amount of principal repaid $P_{i,t}^{M1}$. Principal repayment is computed as the difference between the total regular or typical annual payment the household makes and the implicit interest payments based on the reported interest rate for the loan and the currently outstanding amount. Thus in some cases $P_{i,t}^{M1}$ might be negative which corresponds to a situation in which actual payments are less than the interest due. This means that unpaid interest accumulates and slowly increases the outstanding amount and thus the principal 'repayment' is negative. In both cases the change in the outstanding liability is defined as the negative principal repayment $\Delta D_{i,t}^{M1} = -P_{i,t}^{M1}$.

If the amount initially borrowed and currently outstanding are equal ($t > j$ and $B_{i,j}^{M1} = D_{i,t}^{M1}$: case 2.2) it is interpreted as the household being in a period of no principal repayment as part of an interest-only agreement. The household only pays interest but no principal for a certain period at the beginning of the contract. Thus the outstanding amount on that mortgage did not change in the year of the interview: $\Delta D_{i,t}^{M1} = 0$.

Households which report an amount currently outstanding larger than the amount initially borrowed ($t > j$ and $B_{i,j}^{M1} < D_{i,t}^{M1}$: case 2.3) are assumed to have fallen behind in payments. Due to the accumulated unpaid interest, the outstanding balance increased and now exceeds the amount initially borrowed. Only a few households report such a constellation, as is expected. It is important to note that also households falling into case 2.1 can be behind on payments. In that case however they underwent an earlier period of repayment and struggled with payments only for a relatively short period such that accumulated interest does not exceed earlier repayments. The change in the mortgage balance for households falling into case 2.3 is also defined as the negative principal repayment: $\Delta D_{i,t}^{M1} = -P_{i,t}^{M1}$. In almost all cases households in this category are paying back principal indicating that they are back on some payment plan. This seems convincing because it is unlikely that a lender will tolerate accumulated unpaid interest over extended periods of time while letting the household keep its mortgage and the corresponding house. One expects a lender in such a situation to seek control of the house and recover the money lent. Case 2.3 is also compatible with a reverse mortgage where, especially elderly borrowers, take out a mortgage to finance retirement expenses and pay back the entire mortgage and all accrued interest when the property is sold after their death. Also negative amortization (negAM) contracts under which neither principal nor interest payments are made for a short period, normally not more than 5 years, at the beginning of the mortgage contract are consistent with a payment pattern of case 2.3.

All cases are summarized in Table 3 which also indicates the number of observations falling into each case in the 2004 wave in the third implicate. 1,984 out of 4,519 households reported an outstanding first mortgage on the primary residence in 2004. 1,466 of these fell into case 2.1 and thus form the biggest group. The overwhelming majority of households in this group (1,431) is paying down debt. Thus, even if at the aggregate level household debt increases, it is only a small proportion of households which takes on debt in a given year. While this is not a surprising result one has to keep it in mind for econometric modelling.

The categorization of households along the two steps is visualized in Figure 2 in Appendix II. By applying the logic from the previous example to the second and third mortgage on the primary residence as well as five additional debt categories (mortgages on other residential property, consumer loans, car and vehicle loans, education loans, other loans for property purchase and home improvements), the author constructs a measure of the total change in household i's liabilities (ΔD_{it}) by summing up the changes of the individual categories:

$$\Delta D_{it} = \Delta D_{i,t}^{M} + \Delta D_{i,t}^{OM} + \Delta D_{i,t}^{CL} + \Delta D_{i,t}^{CAR} + \Delta D_{i,t}^{EDU} + \Delta D_{i,t}^{OL} \tag{2}$$

where $\Delta D_{i,t}^{M}$ is the change in all outstanding mortgages on the primary residence (first, second and third: $\Delta D_{i,t}^{M} = \Delta D_{i,t}^{M1} + \Delta D_{i,t}^{M2} + \Delta D_{i,t}^{M3}$), $\Delta D_{i,t}^{OM}$ is the change in outstanding mortgages on other residential properties, $\Delta D_{i,t}^{CL}$ is the change in outstanding unsecured consumer loans, $\Delta D_{i,t}^{CAR}$ is the change in outstanding car and vehicle loans, $\Delta D_{i,t}^{EDU}$ is the change in outstanding education loans and $\Delta D_{i,t}^{OL}$ is the change in other outstanding liabilities ($\Delta D_{i,t}^{OL} = \Delta D_{i,t}^{OLP} + \Delta D_{i,t}^{HI}$). The definitions and the steps undertaken to define the change in outstanding liabilities for these other categories can be found in Appendix B.

Table 3: Changes in the outstanding amount of the first mortgage on the primary residence (ΔD^{M1}) in 2004

case	step 1	step 2	definition	N	interpretation
1.1	t = j ∧ use = ref		ΔD^{M1} = D-B	168	Mortgage M1 taken out to refinance earlier mortgage.
1.2	t = j ∧ use = extr		ΔD^{M1} = ex	19	Mortgage agreement M1 extended to extract equity.
1.3	t = j ∧ use = extr+ref		ΔD^{M1} = ex-(B-D)	39	Mortgage agreement M1 altered to refinance and to extract equity.
1.4	t = j ∧ use = no prior mortgage		ΔD^{M1} = D	109	No prior mortgage thus M1 counts as new debt.
2.1	t > j	B > D	ΔD^{M1} = -P	1,466	Household is repaying M1.
2.2	t > j	B = D	ΔD^{M1} = 0	164	No repayment yet. Probably interest only period.
2.3	t > j	B < D	ΔD^{M1} = -P	19	Fallen behind payment schedule and interest accumulated.
				1,984	

N is the number of households in each case in the 2004 wave in implicate 3. j corresponds to X802, use to X7137, ex to X7138, B to X804, D to X805 and P is defined as P=X808-(X816/10000*X805) where X808 are transformed to annual payments and X813 is used if X808 is not reported. Observations falling into cases 3 and 4 are dropped prior to 2004 because ex is not observed.

4.2 Estimating a Household-Debt-Accumulation-Function

The specification of a reduced form household borrowing equation is based on the predictions from economic theory which were discussed in the introduction. This section provides more detail. Most importantly the variable of interest is the annual change in total outstanding liabilities (ΔD_{it}) and not the total stock of debt (D_{it}). The simple reason is that only the change is directly related to the flows of the current period whereas the stock depends on past decisions which are not observed. The change in liabilities is scaled by current household income (Y_{it}) in order to obtain a measure of borrowing which is comparable across the population. In contrast, using the growth rate of outstanding liabilities looks appealing at first but debt growth rates are not useful indicators of household debt accumulation. Taking out new debt with very little outstanding liabilities results in misleadingly large growth rates even if the actual amount of new debt is quite small in absolute terms or relative to household income. Beyond that, growth rates are not defined for households with no prior liabilities and using a log approximation is highly problematic because it does not work well for large rates[34]. In addition, the amount of already existing liabilities is not a particularly

[34] These are specific problems of household data. Aggregate data at the state or county level are less prone to these issues because the outstanding debt stock within the county or State most likely will be well above 0 and thus growth rates will yield an indicator of household borrowing which is comparable across counties

relevant criterion in order to assess the extent of a household's indebtedness. Instead income represents the flow out of which the household has to repay principal and interest and thus yields a more informative benchmark. This chapter also does not use assets to scale household liabilities because asset prices tend to be volatile and in boom phases they increase in line with liabilities exposing high leverage only after asset prices collapse. For these reasons income is used to scale household borrowing. The exact definitions of all variables used, in line with the notation of the official documentation of the SCF, can be found in Table 1 in Appendix II.

When thinking about determinants of household debt dynamics the first category to consider is household income for obvious reasons: income is the main source of funding for most expenditures especially for consumption expenditures and thus influences to what extent borrowing is needed to achieve desired spending levels. Different economic theories however give different predictions about the impact of income on household borrowing. The life-cycle model for example predicts that households borrow in order to smooth consumption. If low-income households are exposed to irregular but large income shocks compared to high-income households, then low-income households need to borrow more (relative to their income) in order to smooth these shocks. The reason why high-income households might face regular but relatively small income shocks is that volatile profit income plays a larger role for them, while (long term) unemployment is more likely to hit low-income households. In contrast to the life-cycle model, Post Keynesian theory emphasises different propensities to consume and save along the income distribution and thus predicts low-income households to borrow more relative to their income compared to higher income households.

If one thinks about the role of income on household borrowing in a cross section context it becomes clear that there is an important difference between borrowing households and non-borrowing households. Implicitly the previous paragraph focused on borrowing households. However since only a small proportion of households actually borrows in any current year but borrowing results in a lengthy period of repayments, an asymmetry arises between the two groups for two reasons. First, if higher income households borrow less relative to their income, this implies that also their repayments are smaller relative to their income. However because borrowing is defined as an increase in outstanding liabilities ($\Delta D_{it} > 0$) and repayment as a decrease ($\Delta D_{it} < 0$), income will have a negative impact on borrowing but a positive impact on repayments when borrowing as well as repayments are expressed relative to household income. The reason is that a negative sign of the income coefficient in the borrowing sample implies that richer households borrow less relative to their income. In contrast, a positive sign of the income coefficient in the non-borrowing sample,

---

or States. However in case of a high degree of heterogeneity in income dynamics, also growth rates at the aggregate level might be a misleading indicator.

implies that richer households repay less because a positive sign implies less negative (i.e. smaller) repayments. Due to these different signs, analysing borrowing and non-borrowing households together in one same sample will yield meaningless averages. The second reason why an asymmetry between borrowing and non-borrowing household emerges is that there is no reason to believe that household characteristics such as income will influence borrowing and repayment decisions in a symmetric way. For example if high income households choose to have shorter repayment periods one can easily think of examples where the proportion of borrowing relative to income between high and low income households and the proportion of repayments (relative to income) between high and low income households is different. Therefore this chapter will investigate the two groups separately.

Beyond income the model incorporates measures of financial wealth (FW_{it}) and most importantly residential real estate assets[35]. The actual purchase of real estate is one of the most important reasons why households take on debt. For that reason this paper distinguishes between housing assets which existed before the beginning of the period and the value of residential real estate purchased in the year of the interview. Thus the measure of housing wealth used in the paper (HW_{it}) consists of the current value of the primary residence and any other residential real estate minus the value and any potential capital gains on residential real estate purchased in the current period. The value at the time of purchase (thus excluding potential capital gains within the year) of residential real estate obtained in the current period (REP_{it}) enters the empirical model separately. For exact definitions readers are referred to Table 1 in Appendix II. Being able to include detailed information on asset purchases in addition to pre-period stocks is a major advantage of SCF data when it comes to assessing the explanatory power of the housing hypothesis. Since asset purchases are the most important reason to take on debt for the majority of households not being able to include this kind of information in one's analysis will result in severely biased estimates. Economic theory predicts financial as well as real estate assets to influence household borrowing due to the presence of credit constraints and 'pure' wealth effects. Thus for both measures one expects a positive cross section effect on borrowing relative to household income. Due to the negative sign of liability changes in the non-borrowing sample, the effect of housing wealth is expected to be negative in the non-borrowing sample while the effect of current real estate purchases is expected to be close to zero since there are simply not many households being able to purchase without taking on debt.

Since the aim of the chapter is to assess the role of relative income on household borrowing, the regression model also includes various measures of peer group income ($\tilde{Y}_{it}$). Several definitions are

[35] This means real estate purchased for investment purposes (buy-to-let) is not part of this housing wealth measure because the SCF does not report debt related to these properties separately but reports net values.

used and the details are discussed in the subsection below. For now it is sufficient to state that under the ECH one expects those households being exposed to higher levels of peer group income, while holding their own income constant, to borrow more in order to keep up with the expenditures of that peer group and thus a positive effect. For the group of non-borrowing households one expects their repayments to be higher if they borrowed more in the past and thus a negative effect. Thus for similar reasons as in the case of income and housing wealth, most likely there is an asymmetry between borrowers and non-borrowers with respect to the expected signs.

The ISH hypothesis rests on the argument that households borrowed in order to sustain consumption levels in a situation of declining incomes. The argument is that generally people are unwilling to cut once-achieved standards of living. It is much harder to adjust downwards than upwards. Thus the regression also controls for those households whose income is lower than in a normal year by means of an indicator variable ($dlinc_{it}$). If past expenditure levels are important for current spending decisions those households with abnormally low incomes should borrow more or repay slower. Another interpretation is that households with abnormally low incomes will borrow less (repay quicker) out of a precautionary motive. Besides the relative position within the income distribution Duesenberry stressed past consumption as an important reference point (Duesenberry 1949).

Finally, outstanding liabilities at the end of the previous period (D_{it-1}) are part of the empirical model. Since for most households borrowing results in a lengthy period of repayments, the amount still outstanding is an important predictor of these payments. On the other hand, for borrowing households one expects that already highly indebted households are less likely to be able to borrow even more. One potential exception arises if one thinks about households as 'Minskian agents' whose finance structure becomes more and more risky over the boom period, eventually ending up in what Minsky labelled *Ponzi finance* which describes agents not able to repay principal nor interest out of current cash flows and who need to borrow for these payments. This approach predicts higher indebted households to borrow even more. Most likely however such behaviour is observed only over short periods of time and not over 13 years as in the current sample. Based on these considerations the regression is specified in the following way:

$$\begin{aligned}\frac{\Delta D_{it}}{Y_{it}} = \alpha^S + \beta_1^S \operatorname{ihs}(Y_{it}) + \beta_2^S \operatorname{ihs}(\tilde{Y}_{it}) + \beta_3^S \operatorname{ihs}(HW_{it}) + \beta_4^S \operatorname{ihs}(REP_{it}) + \\ \beta_5^S \operatorname{ihs}(D_{it-1}) + \beta_6^S \operatorname{ihs}(FW_{it}) + \beta_7^S dlinc_{it} + \beta_8^S dyear_t + \boldsymbol{\beta}^S \boldsymbol{X}_t + \varepsilon_{it}\end{aligned} \tag{3}$$

where $S = \{B, NB\}$ indicates the subsamples of borrowing and non-borrowing households, $dyear_t$ is a set of year dummies and $\boldsymbol{X}_t$ is a matrix containing household characteristics such as age and age-squared of the household head as well as dummies for the presence of children, for being married or living with a partner, for being part of the labour-force, for having a college degree, for having been denied when applying for credit and a set of race dummies. The motivation of including such a rich set of household characteristics is to control for household heterogeneity.

The set of time dummies as well as the indicator variable for households having been turned down when applying for credit over the last four years are a way to control for shifts in credit supply conditions. If lenders are more willing to hand out larger loans relative to a households' income level or stock of assets then this should affect households homogeneously across time and thus allowing for time varying intercepts should capture that. In contrast, if lenders are become increasingly willing to accept customers who were not able to obtain credit in the past, then the effect of such a shift in credit conditions should affect those households which were denied access to credit in the past. A decline in the proportion of households which have not been able to obtain credit in the past then indicates a shift in the supply of credit towards low quality borrowers.

Equation (3) is estimated by pooled OLS using probability weights. Weighted estimation is important because due to oversampling, households from the upper tail of the income and wealth distribution would be overrepresented if all observations were implicitly assigned equal weights as in case of unweighted estimation. Standard errors are based on a bootstrap procedure which re-estimates each regression 999 times using a set of replicate weights instead of the initial weights. Replicate weights are part of the SCF dataset and are designed to replicate the sampling process. Since the Federal Reserve Board does not publish the details of their sampling procedure in order to protect the privacy of the survey participants the replicate weights are the only way to obtain standard errors which take stratification and oversampling properly into account. Instead of taking logarithms of variables in Dollar terms, they are transformed using the (unscaled) inverse hyperbolic sine transformation (ihs) which is defined for zero and negative valued observations[36] (MacKinnon & Magee 1990; Friedline et al. 2015).

4.3 Defining Reference Income

Even though the idea of the ECH that an increasingly polarized income distribution triggers debt financed spending as left-behind households attempt to maintain their perceived social status is quite simple, the task of testing it empirically is not. The main difficulty lies in defining adequate reference groups.

Before going into details it is important to emphasise again that the expenditure cascades hypothesis strictly relies on upward looking status comparison. Only if households compare themselves with others higher up in the income distribution will an increase in income inequality

[36] The inverse hyperbolic sine transformation is defined as: $ihs(x) = \ln(x + \sqrt{x^2 + 1})$. The attractiveness of this transformation stems from the fact that it can be applied to zero and negative values while the interpretation of ihs transformed data in a regression context is equivalent to log-transformed data since $\frac{\partial ihs(x)}{\partial x} = \frac{1}{\sqrt{x^2+1}}$ which asymptotically approaches $\frac{1}{x}$ as x increases.

lead to a cascade of debt-financed expenditures. In contrast, if households rely on average income as their point of reference, a polarisation in the distribution of income would not lead to an increase in spending and borrowing for those households with above-average incomes. Even though the mean is a commonly used point of reference in the literature (Alpizar et al. 2005; Maurer & Meier 2008; Alvarez-Cuadrado & Vilalta 2012; Alvarez-Cuadrado et al. 2012), it is not consistent with the ECH.

Instead of mean income, the following three definitions of reference income, motivated by upward looking status comparison, will be used. The first definition assumes that households compare themselves to richer peers at the top of the income distribution. In particular it is assumed that the p^{th} income percentile is the point of reference for all households below that percentile. This approach is similar to Bertrand and Morse (2013) who use the 80^{th} and 90^{th} percentile of the income distribution within the state household i lives in as the cut-off. While the assumption that status comparison takes place within regionally defined communities is plausible, US States are too large to serve as realistic proxies for such communities. The previous chapter's approach to reference groups is adopted. They are defined based on educational achievement (college or less-than-college) and the racial background (white or black) of households, resulting in 4 groups[37]. The rationale for grouping along educational achievement and race is that both variables are important factors in defining the social sphere in which individuals engage with others through work, residency and leisure activities. The SCF provides four racial categories: white, black, hispanic and other. The reason for excluding the other category in the definition of reference groups is that it is a residual category which does not identify a homogeneous group. However reference groups are motivated by the idea that households engage with and compare themselves to similar peers. The reason for dropping the hispanic category is simple: The number of observations within the college-hispanic cells are too small to reliably calculate within group income distribution measures[38].

In contrast to Bertrand and Morse (2013) not only the 80^{th} and 90^{th} percentiles are used as cut-off points but the percentiles $q = \{99, 95, 90, 80\}$. Since mainly the top 5% of households enjoyed above average income growth in the period 1995 to 2007, including the 99^{th} and 95^{th} percentiles is crucial in analysing the impact of income inequality on household spending and borrowing[39]. This

[37] The SCF includes identifiers for 9 regions, based on groups of US States, in the years 1995 and 1998 but not from 1998 onwards. As a robustness check income percentiles were computed within these regions and compared to the results based on education and race groupings, see Table A4 in the Appendix. Results do not differ systematically between specifications relying on geographically identified groups and race-education identified groups.

[38] For example in 1995 there are only 31 observations in the college-hispanic cell compared to 71 in the college-black cell, 1,602 in college-white, 141 in no-college-hispanic, 285 in no-college-black and 1,839 in no-college-white.

[39] This becomes even more important when grouping households along education and race as income gains are even more concentrated towards the top within these finer groups compared to the nationwide income distribution.

first definition of reference income based on the percentile cut-off is denoted $\tilde{Y}_{it}^{pq}$ where p indicates the use of percentiles as the point of reference and q indicates which percentile was used as cut-off. With reference income defined in that way one expects that households which are exposed to higher levels of top-incomes in their education-race group borrow more in order to finance a similar expenditure level as these richer peers. In order to distinguish peer effects from the effects of education and race a full set of race-education dummies along with all interactions is included in the regression. Thus the effect of $\tilde{Y}_{it}^{pq}$ is identified by the variation over time only. In a pure cross section $\tilde{Y}_{it}^{pq}$ would not be distinguishable from the effects of education and race captured by the set of dummies and their interactions.

The second definition of reference income is very similar but instead of using income percentiles as reference points, the average household income above a chosen percentile is used as the point of reference. Compared to the first definition, the average income above a certain percentile also reveals information about the households above that percentile. In particular the evolution of the average income of the top 5% of all households (i.e. households above the 95th percentile) for example, is a better indicator about the income share of that group and thus the evolution of the income distribution than just the 95th percentile. Percentiles $q = \{99, 95, 90, 80\}$ are used as cut-off points. This second definition of reference income based on averages above a certain percentile is labelled $\tilde{Y}_{it}^{avq}$ where avq indicates the use of average income above percentile q. Since for those households with income levels in excess of the cut-off there is no reference point defined, these are excluded from the estimation. This is the case for $\tilde{Y}_{it}^{avq}$ as well as $\tilde{Y}_{it}^{pq}$. Under the ECH the effect of $\tilde{Y}_{it}^{avq}$ on household borrowing is expected to be positive.

The third definition is closely related to the measure of relative income used by Carr and Jayadev (2015). They cluster households based on which US State they live in and then the proportion of households within that group which are richer than household i is used as a proxy for household i's relative income. Since US States are too large to serve as meaningful proxies for communities within which status comparison takes place, this paper again relies on education and race based groups g:

$$\tilde{Y}_{it}^{head} = \frac{\sum I[I = 1|Y_{-igt} > Y_{igt}]}{N_{gt}} \tag{4}$$

where $I \in \{0,1\}$ is an indicator variable equal to 1 when a given observation's income is greater than observation i's income. $g \in \{1, \dots, 4\}$ represents the four groups of households clustered by education (college or less-than-college) and race (white or black). Thus $\tilde{Y}_{it}^{head}$ corresponds to the tail distribution[40] of income within group g. As such it is very different from the previous two definitions because it does not provide an absolute measure of within-group income concentration which is comparable across groups. Instead it provides information of the relative position for individual households within the group. $\tilde{Y}_{it}^{head}$ is expected to yield a positive effect on household debt accumulation in the borrowing and in the non-borrowing sample. Higher values of $\tilde{Y}_{it}^{head}$ are associated with households which are relatively poor in their race-education group, and are therefore exposed to a large number of richer peers and thus are expected to borrow more or repay more slowly. There are three important drawbacks with this definition of relative income. First, it implies that expenditure cascades become stronger towards the bottom of the within group income distribution. A priori it is not clear whether this should be the case. Instead if households are only partially able to keep up with their richer peers it might even be the case that expenditure cascades decline towards the bottom of the within group distribution. Second, defining relative income in that way does not provide any information about the degree of within group income concentration towards the top, which is central for the expenditure cascades argument. Thus using $\tilde{Y}_{it}^{head}$ is equivalent to focusing on the within group position in the income distribution while ignoring how the absolute degree of inequality within that group compares across groups. Put differently, one assumes that households do not care by how much their peers are richer but care only about how many other households are richer. $\tilde{Y}_{it}^{head}$ is essentially a headcount of richer households. The ECH however relies on a situation of growing within group income inequality. Third, this concept does not measure how the distribution of income changed over time. It can only be used to explain differences between households.

Summing up: the chapter uses three different versions of relative income: The first two measure the absolute degree of inequality within education-race groups and are defined as the q^{th} percentile of the education-race group income distribution ($\tilde{Y}_{it}^{pq}$) and the average income above the q^{th} percentile ($\tilde{Y}_{it}^{avq}$). The third measure is defined as the tail distribution of income within education-race groups ($\tilde{Y}_{it}^{head}$) which corresponds to a head count of households richer than household i.

4.4 Determinants of US Household Borrowing

Table 4 presents the first set of results. Equation (3) is estimated for three different samples: the borrowing and non-borrowing subsamples as well as the full sample (containing borrowing and non-borrowing households). Since all independent variables denominated in monetary terms are subject to the inverse-hyperbolic-sine transformation the estimated coefficients in Table 4 can be interpreted as changes in household borrowing in % of income due to a 1% increase of the corresponding explanatory variable. In order to interpret the results correctly it is important to bear

[40] Tail distribution or the complementary cumulative distribution function which is defined as 1 minus the cumulative distribution function.

in mind that repayment is ‘negative borrowing’ and a positive coefficient in the non-borrowing sample indicates less ‘negative borrowing’ and thus lower repayments. As expected, there is a pronounced asymmetry in household behaviour between borrowing and non-borrowing groups as is indicated by the vastly different coefficients reported in columns (1) and (2). Combining borrowing and non-borrowing households in one group as in column (3) yields misleading averages which fail to adequately describe the behaviour of either group. For this reason, the borrowing and non-borrowing sample are analysed separately from here on.

Beyond these differences in behaviour, results in Table 4 indicate that with rising income levels, households borrow less relative to their income (columns 1 and 4) and also use a smaller proportion of their income for repayments (columns 2 and 5). Thus lower income households leverage up more, relative to income. These findings are compatible with the life-cycle model if one assumes that higher income households face smaller shocks relative to their income. It is also compatible with the Post Keynesian stock-flow-consistent modelling tradition which assumes borrowing for consumption purposes will be smaller relative to income for high income households due to lower marginal propensities to consume. Also the ECH predicts a negative impact of income on household borrowing because richer households are less dependent on borrowing in order to finance a given level of status consumption.

Table 4 reveals that the coefficients on housing wealth (HW) and real estate purchases (REP) are positive and highly statistically significant (columns 1 and 4). It indicates that holding everything else constant higher levels of housing wealth lead to more borrowing. In addition, purchasing residential real estate in the year of the interview is associated with a significantly higher level of borrowing. Since purchasing real estate without taking out a mortgage is highly unusual the insignificant coefficient in the non-borrowing sample is not surprising. Overall these results are very well in line with a life-cycle model incorporating credit constraints, with a Minskian interpretation of the relationship between assets and leverage and also with the notion of households being anchored by assets-to-income ratios as argued by the stock-flow-consistent modelling tradition. Thus, the importance of assets and especially housing assets as predicted by the HH, is strongly supported by the data.

The coefficient on financial wealth (FW) is neither statistically significant in the borrowing nor in non-borrowing sample. This might be due to the fact that even though financial wealth should ease credit constraints in a similar way as housing wealth does, banks are much more reluctant to accept financial assets with potentially highly volatile prices as collateral. The coefficient on total liabilities at the beginning of the period (D_{it-1}) is only statistically significant in the non-borrowing specification indicating that higher indebted households use a larger proportion of their income for repayments. The lack of a significant coefficient on outstanding liabilities in the borrowing sample is unexpected since the life-cycle model which emphasizes net-wealth as well as the stock-flow-consistent modelling tradition predict a negative effect. On the other hand a Minsky inspired interpretation suggests that heavily indebted households might have to rely on new borrowing in order to keep up with payments, predicting a positive effect. If both arguments are valid potentially at different stages of the credit cycle (i.e. the Minsky argument only holding shortly before the peak) they might cancel each other out over longer sample periods leaving a statistically insignificant result.

While it became clear that assets and outstanding liabilities are crucial in determining household borrowing decisions, Table 4 also reveals the importance of the income distribution. The coefficient on reference income, defined as the average income of the top 1% of households within the education and race groups ($\tilde{Y}_{it}^{av99}$) has a highly significant positive coefficient in the specification of borrowing households (column 1). Thus indicates that a 1% increase in average incomes at the top leads to an increase in household borrowing by 0.2% of household income. This result strongly supports the ECH: Households being exposed to a more unequal distribution of income borrow more relative to their household income. For the non-borrowing sample the coefficient is not statistically different from zero and thus it seems that being exposed to richer peers does not slow down repayment efforts. The fact that repayment conditions are agreed on in advance provides an explanation. Put differently, households have a hard time explaining to their bank a delay in payments due to their neighbour buying a bigger car.

In specifications (1) to (3) in Table 4, the coefficient on average top incomes ($\tilde{Y}_{it}^{av99}$) does not depend on any measure of credit constraints. However it is not clear how households without any relevant assets should be able to secure additional borrowing in order to finance status driven expenditures. Thus it might be the case that only unconstrained households are able to engage in status seeking borrowing cascades. In order to control for such a scenario, columns (4) and (5) of Table 4 report the results of estimating equation (3) with the measure of average top income ($\tilde{Y}_{it}^{av99}$) interacted with a dummy for residential real estate ownership ($HW1$ indicating ownership and $HW0$ no ownership). The results do indeed support the hypothesis that reference incomes only affect non-credit constrained households, in this case owners, since the coefficient on average top incomes is not statistically not different from 0 for non-owners. At the same time the coefficients on the other variables remain unchanged compared to the specification without the interaction terms. This is an important finding because it indicates that while the distribution of income does matter via status comparison across households and so-called expenditure cascades it also indicates that only those households which possess some form of collateral are actually able to react to this kind of peer pressure by increasing their own expenditures using borrowed money. So it seems that household borrowing is determined in a complex interaction of different channels and it seems that in isolation neither the expenditure cascades hypothesis nor the hypothesis of

Minskyian households are able to fully explain household borrowing[41]. Only an interaction of inequality and asset prices and thus a conditional version of the ECH can provide such an explanation.

Table 4: Baseline Specification

	(1)	(2)	(3)	(4)	(5)
sample	ΔD/Y>0	ΔD/Y≤0	full	ΔD/Y>0	ΔD/Y≤0
$\tilde{Y}$ cut-off	99th perc	99th perc	99th perc	99th perc	99th perc
$\tilde{Y}$ definition	av. inc.	av. inc.	av. inc.	av. inc.	av. inc.
Y	-0.314	0.042	0.001	-0.314	0.042
	(0.024)***	(0.003)***	(0.005)	(0.023)***	(0.003)***
$\tilde{Y}^{av99}$	0.216	0.009	0.056		
	(0.080)***	(0.006)	(0.019)***		
$HW0\#\tilde{Y}^{av99}$				0.089	0.004
				(0.115)	(0.006)
$HW1\#\tilde{Y}^{av99}$				0.289	0.010
				(0.085)***	(0.007)
HW	0.086	-0.002	0.006	0.090	-0.003
	(0.017)***	(0.001)	(0.003)*	(0.017)***	(0.001)**
REP	0.659	0.011	0.539	0.658	0.012
	(0.047)***	(0.011)	(0.038)***	(0.048)***	(0.011)
FW	0.008	-0.001	0.000	0.008	-0.001
	(0.006)	(0.001)	(0.001)	(0.006)	(0.001)
D_{t-1}	-0.005	-0.024	-0.024	-0.007	-0.024
	(0.009)	(0.001)***	(0.002)***	(0.009)	(0.001)***
$constant$	0.758	-0.555	-0.669	2.607	-0.480
	(1.176)	(0.086)***	(0.269)**	(1.710)	(0.093)***
N	2,229	14,270	16,510	2,229	14,270
av. R^2	0.65	0.27	0.51	0.65	0.27
time effects	yes	yes	yes	yes	yes
household characteristics	yes	yes	yes	yes	yes

Dependent variable: ΔD/Y. All independent variables are subject to the inverse hyperbolic sine transformation. Y~ is defined as the average income above the 99th percentile within education-race groups (college/less-than-college and black/white). Coefficients are estimated by OLS using probability weights. Bootstrapped standard errors are obtained by re-estimating the regression 999 times using a set of 999 replicate weights. The R^2 is the average across all implicates. Stars indicate 1% (***), 5% (**) and 10% (*) significance levels. A full table including the missing household characteristics and time effects is provided in Table 2 in Appendix II.

The results so far are based on a specification using $\tilde{Y}_{it}^{av99}$ as the reference income definition. In order to check the robustness of these results Table 5 reports additional specifications incorporating alternative definitions of reference income. Using other forms of reference income than $\tilde{Y}_{it}^{av99}$ serves two purposes. First, using cut-off points other than the top 1% serves as a robustness check to the question whether expenditure cascades are triggered by concentration of

[41] Table 3 in Appendix II presents fully separated specifications for owners and non-owners.

income at the very top of the distribution. Second, using relative income measures based on percentiles ($\tilde{Y}_{it}^{pq}$) instead of averages as well as the relative position within the group ($\tilde{Y}_{it}^{head}$) provides a test of other mechanisms through which status comparison might take place and expenditure cascades might occur.

Table 5 demonstrates that using different percentiles as cut-off points for computing the average income of households above the qth percentile still yields statistically significant and positive relative-income-coefficients (column 1: 95th percentile and column 2: 80th percentile) while the coefficients on the other variables remain qualitatively unchanged. However only the 99th percentile cut-off used in Table 5, yields a statistically significant coefficient at the 1 % level, indicating that expenditure cascades are triggered from the very top of the income distribution. When only the income at the cut-off is used in estimations (columns 3 to 5 and 9 to 11) instead of the average income above the cut-off, there is no statistically significant reference income coefficient. Defining relative income as the proportion of richer households ($\tilde{Y}_{it}^{head}$) yields a negative and statistically highly significant estimate (column 6) both for owners and non-owners. Thus it seems that households towards the bottom of the within group income distribution borrow less even when the level of income is controlled for separately. The negative coefficient on $\tilde{Y}_{it}^{head}$ in the non-borrowing sample (column 12) indicates that households towards the bottom of the within group income distribution use a larger proportion of their income for repayments. Thus the results for $\tilde{Y}_{it}^{head}$ do not support the expenditure cascades hypothesis. Due to the focus on within group variation of $\tilde{Y}_{it}^{head}$ instead of across group variation, this paper prefers the other two definitions of relative income based on top group averages ($\tilde{Y}_{it}^{avq}$) and percentile cut-off values ($\tilde{Y}_{it}^{pq}$).

The coefficients on housing wealth (HW) and real estate purchases in the current year (REP) do not change in a qualitative way when using different measures of reference income. Thus Table 3 demonstrates the robust support of the data for the HH. However up to now the discussion of the results solely focussed on the statistical significance and the signs of the estimated coefficients in order to judge whether they are in line with predictions from economic theory. The next step is to assess the economic significance of the estimated model and to compare the predictive power of the individual variables. This is done in the next section.

Table 5: Additional Reference Income Definitions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
sample	ΔD/Y>0	ΔD/Y>0	ΔD/Y>0	ΔD/Y>0	ΔD/Y>0	ΔD/Y>0	ΔD/Y≤0	ΔD/Y≤0	ΔD/Y≤0	ΔD/Y≤0	ΔD/Y≤0	ΔD/Y≤0
$\tilde{Y}$ definition	$\tilde{Y}^{av95}$	$\tilde{Y}^{av80}$	$\tilde{Y}^{p99}$	$\tilde{Y}^{p95}$	$\tilde{Y}^{p80}$	$\tilde{Y}^{head}$	$\tilde{Y}^{av95}$	$\tilde{Y}^{av80}$	$\tilde{Y}^{p99}$	$\tilde{Y}^{p95}$	$\tilde{Y}^{p80}$	$\tilde{Y}^{head}$
$\tilde{Y}$ cut-off	95th perc	80th perc	99th perc	95th perc	80th perc	none	95th perc	80th perc	99th perc	95th perc	80th perc	none
Y	-0.307	-0.327	-0.313	-0.307	-0.327	-0.517	0.043	0.043	0.042	0.043	0.043	0.025
	(0.026)***	(0.032)***	(0.023)***	(0.026)***	(0.032)***	(0.073)***	(0.003)***	(0.004)***	(0.003)***	(0.003)***	(0.004)***	(0.005)***
$HW0\#\tilde{Y}$	0.121	0.191	0.047	0.159	0.294	-0.645	-0.004	-0.008	-0.010	-0.012	0.005	0.002
	(0.171)	(0.235)	(0.141)	(0.245)	(0.380)	(0.215)***	(0.010)	(0.014)	(0.009)	(0.013)	(0.027)	(0.015)
$HW1\#\tilde{Y}$	0.203	0.330	0.014	0.052	0.467	-0.666	0.009	0.017	0.004	0.011	0.029	-0.089
	(0.105)*	(0.142)**	(0.071)	(0.164)	(0.305)	(0.197)***	(0.007)	(0.012)	(0.005)	(0.010)	(0.019)	(0.013)***
HW	0.110	0.115	0.095	0.113	0.115	0.095	-0.004	-0.004	-0.003	-0.004	-0.004	-0.008
	(0.018)***	(0.019)***	(0.017)***	(0.018)***	(0.019)***	(0.016)***	(0.001)***	(0.002)***	(0.001)**	(0.001)***	(0.002)***	(0.001)***
REP	0.752	0.857	0.659	0.751	0.856	0.606	0.014	0.017	0.012	0.014	0.017	0.012
	(0.052)***	(0.052)***	(0.047)***	(0.051)***	(0.052)***	(0.045)***	(0.012)	(0.016)	(0.011)	(0.012)	(0.016)	(0.010)
FW	0.007	0.006	0.008	0.007	0.006	0.005	-0.001	-0.001	-0.001	-0.001	-0.001	-0.001
	(0.006)	(0.007)	(0.006)	(0.006)	(0.007)	(0.006)	(0.001)	(0.001)*	(0.001)	(0.001)	(0.001)*	(0.001)*
D_{t-1}	-0.005	-0.004	-0.007	-0.005	-0.004	-0.009	-0.025	-0.028	-0.024	-0.025	-0.027	-0.024
	(0.009)	(0.010)	(0.009)	(0.010)	(0.010)	(0.009)	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***	(0.001)***
$constant$	2.238	1.669	3.258	1.839	0.538	6.511	-0.383	-0.331	-0.297	-0.288	-0.496	-0.251
	(2.345)	(3.025)	(1.921)*	(3.127)	(4.612)	(0.945)***	(0.127)***	(0.184)*	(0.122)**	(0.163)*	(0.331)	(0.065)***
N	2,242	2,024	2,477	2,242	2,024	2,722	14,183	12,971	15,931	14,183	12,971	18,358

Dependent variable: ΔD/Y. All $ valued independent variables are subject to the inverse hyperbolic sine transformation. Coefficients are estimated by OLS using probability weights. Bootstrapped standard errors are obtained by re-estimating the regression 999 times using a set of 999 replicate weights. Stars indicate 1% (***), 5% (**) and 10% (*) significance levels.

4.5 Economic Significance

In order to compare the explanatory power of the ECH and HH, contributions to the predicted value of household borrowing are calculated. For example in the case of housing wealth (HW) the contribution of housing wealth to the accumulation of debt in year t aggregated over all households is calculated as:

$$\Omega_t^{HW,S} = \sum_{i=1}^{N_s} \hat{\beta}_3^S \, ihs(HW_{it}) \, Y_{it} \tag{5}$$

where $\hat{\beta}_3^S$ is the estimated coefficient of housing in equation (3) and $S \in \{B, NB\}$ indicates that the computation is done for the borrowing (B) as well as the non-borrowing (NB) subsample. Equivalent computations are carried out for all independent variables in the model.

Table 6 presents the contributions of the main variables of interest: household income, relative income and household balance sheet items. The overall model prediction is also reported in the first column. Based on equation (5) contributions are computed for each subgroup, expressed as the difference relative to 1995 and scaled by aggregate income. For example, the contribution of housing wealth reported in Table 6 is computed as:

$$\Delta\Omega_t^{HW} = (\Omega_t^{HW,B} + \Omega_t^{HW,NB} - \Omega_{1995}^{HW,B} - \Omega_{1995}^{HW,NB})/\boldsymbol{Y}_t \tag{6}$$

where $\boldsymbol{Y}_t$ is aggregate income of the actual sample, excluding households which identify as hispanic or other and households with borrowing to income ratios smaller than -250% and bigger than 500% but including those households above the cut-off points used for the definition of relative income. Panel A of Table 6 uses the average income of households above the 99th percentile of the within group income distribution as reference income (coefficients taken from Table 4, columns 4 and 5). Panel B uses the average income of households above the 95th percentile of the within group income distribution (coefficients taken from Table 5, columns 1 and 7). The different measures of reference income are the main reason why contributions of individual variables as well as predicted total borrowing differ across the two Panels: Panel B is based on a smaller sample excluding not only the top 1% of the within group income distributions but the top 5%. However the sample size is not the only reason for the variation in the results. Another important reason is that top 1% income shares and thus average income increased much stronger than top 5% income shares over the sample period.

Table 6: Contributions to household borrowing relative to 1995

Panel A: top 1%

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	total	$HW0 \# \tilde{Y}^{av99}$	$\tilde{Y}^{av99} + HW$	Y	REP	FW	D_{t-1}
1998	0.8%	-0.5%	5.9%	-3.3%	0.6%	0.7%	-1.3%
2001	1.1%	-0.1%	6.2%	-2.3%	1.1%	0.4%	-3.2%
2004	7.0%	1.2%	11.2%	-8.6%	6.8%	1.7%	-3.6%
2007	2.3%	-0.3%	8.7%	-2.9%	3.5%	0.5%	-4.7%

Panel B: top 5%

	total	$HW0 \# \tilde{Y}^{av95}$	$\tilde{Y}^{av95} + HW$	Y	REP	FW	D_{t-1}
1998	1.0%	-0.8%	3.7%	-1.3%	0.3%	0.3%	-1.0%
2001	0.8%	-0.6%	4.3%	-0.7%	0.3%	0.3%	-2.3%
2004	5.6%	1.4%	6.7%	-4.1%	4.6%	0.8%	-3.0%
2007	2.0%	-0.4%	4.6%	0.1%	2.4%	0.2%	-3.5%

Contributions to household borrowing are computed based on results from columns 4 and 5 of Table 4 for Panel A and from columns 1 and 7 of Table 5 for Panel B using equation (6). Values correspond to changes in household borrowing with respect to 1995 expressed in % of aggregate sample income (1998: $6,893 bn; 2001: $8,611 bn; 2004: $8,584 bn; 2007: $9,573 bn in 2013 prices). Total refers to the total change in household borrowing. Total is based on all variables and thus is not equivalent to the sum of the variables displayed in the Table. Full tables containing the contributions of all variables in the model can be found in Appendix II.

Since the reference income coefficient is interacted with a housing ownership dummy in the specification used to calculate the contributions, the contribution of changes in housing wealth and reference income cannot be distinguished for the group of owners[42]. For that reason Table 6 presents the contribution of reference income for non-owners (column 2) and a combined contribution of reference income and real estate wealth for owners (column 3). First, Table 4 shows that household borrowing strongly increased in 2004: the difference in household borrowing between 2004 and 1995 amounts to 7% of household income or $600 billion in 2013 prices[43]. Rsing within group top incomes contributed to additional household borrowing of 1.2% ($104 billion with $\tilde{Y}^{av99}$) and 1.4% ($120 billion with $\tilde{Y}^{av95}$) of household income among non-owners, depending on the relative income measure used. In contrast, increasing housing wealth and top incomes stimulated household borrowing during the same period by 11.2% ($963 billion with $\tilde{Y}^{av99}$) and 6.7% ($575 billion with $\tilde{Y}^{av95}$) of household income. With both specifications, the partial contributions exceed the total amount of additional borrowing in all years, emphasising the importance of the interaction of housing and relative income dynamics. In addition to existing assets, real estate purchases in the current period also contributed significantly to the leveraging up process of the household sector: by 6.8% ($584 billion with $\tilde{Y}^{av99}$) and 4.6% ($395 billion with

[42] The reason is that the positive effects of reference income and housing wealth need to be summed up with the negative effect of the ownership dummy and there is no meaningful way of separating the effect of reference income and ownership and changes in housing wealth.
[43] All subsequent dollar values are expressed in 2013 prices.

$\tilde{Y}^{av95}$) respectively. The growth in household income reduces household borrowing over the same period by 8.6% (\$740 billion with $\tilde{Y}^{av99}$) and 4.1% (\$352 billion with $\tilde{Y}^{av95}$). Table 6 also demonstrates the crucial role of past liabilities for household borrowing (-3.6% and -3% in 2004). Increasing debt levels slowed down the accumulation of new liabilities. The negative contribution of past liabilities on borrowing almost exclusively works through the group of non-borrowing households: With increasing debt levels households have to use larger proportions of their disposable income to repay principal and interest.

Based on these numbers, this chapter confirms the empirical support for the HH which was already established by the descriptive analysis in the previous chapter. The regression results show that expenditures for house purchases as well as the value of residential real estate explain a large part of the increase in household indebtedness prior to 2007. However there is also conditional evidence for the ECH: homeowners exposed to higher within group inequality tend to borrow more relative to their disposable income. This latter result shows the fundamental role of real estate assets in easing credit constraints. Expenditure cascades only materialise if households can rely on collateral for borrowing in order to keep up with their richer peers. Thus simple statements of the expenditure cascades hypothesis which ignore the limiting role of credit constraints in general and the role of real estate in particular do not describe the pre-2007 experience of the US very well. The interaction between homeownership and income inequality calls for an integrated analysis of household borrowing which takes relative income as well as household balance sheet dynamics properly into account, which is missing from the current literature.

4.6 Summary of Chapter 4

The formal regression analysis confirms the key result from the descriptive analysis: housing assets and real estate purchases are the dominant factor in explaining household borrowing behaviour. Support for the HH emerges in a strong and very robust way from these results. This strengthens the conclusion from the previous chapter: Any analysis of US household sector debt needs to take the housing market into account because property prices and mortgage borrowing are the most important drivers of household indebtedness. Ignoring them has the potential to yield seriously biased results.

The fact that household borrowing is higher for homeowners who are exposed to higher within group inequality compared to non-homeowners and homeowners not exposed to high inequality is consistent with the ECH, conditional on homeownership. The finding that the statistical significance of the direct relationship between within group income inequality and household borrowing breaks down when using the richest 10% or 20% of households instead of the richest 1%, can be interpreted as evidence that expenditure cascades are triggered by conspicuous consumption at the very top of the within group income distribution. The interaction between homeownership and within group inequality is interpreted as evidence for the important role collateral in plays for household borrowing. Credit constraints are binding in the sense that even if households want to take on debt to keep up with the Joneses, their bank will only allow them to do so if they already have some form of collateral or use the loan in order to buy something which can be used as collateral. Overall, consistent with the previous chapter, the results strongly support the asset focused few of household borrowing (Bezemer & Zhang 2014; Borio 2014; Dynan & Kohn 2007; Mian & Sufi 2011) summarised as the housing hypothesis.

Borrowing as a reaction to falling income does not seem to have played a major role, after controlling for house purchases, the distribution of income and the value of assets separately. This does not mean that those households who faced below normal income levels did not borrow substantial sums, but it seems that also for those households asset price and distributional dynamics explain most of the variation in borrowing behaviour. This means the regression results do not support the ISH.

Alltogether the results presented in this chapter strongly suppor the housing hypothesis which in turn is consisten with different theoretical explanations. First, a life-cycle model with credit constraints is consistent with large effects of real estate asset values on household borrowing. Second, if household expenditures are anchored by stock-flow-norms, then increasing asset valuations will also increase household borrowing. Third, from a Minskian perspective rising asset prices will increase households' confidence and their willingness to take on debt, resulting in higher debt to income ratios. There is some evidence in support of the expenditure cascades hypothesis but only as an explanation of the behaviour of the subgroup of black households. The ECH rests on other-regarding social norms, a concept which is well established in behavioural economics. The ISH, which rests on self-regardind social norms, is not supported by the data. Furthermore there is no evidence that supports the prediction of the life-cycle model that household age is a good predictor of household debt.

Comparing these results with the existing (microeconometric) literature on household borrowing is not straight forward because the regression specifications, concepts and data used are quite different. For example Carr and Jayadev (2015) use panel data and find that while controlling for household income, low-income households borrow more compared to high income households (i.e. a positive coefficient on $\tilde{Y}_{it}^{head}$). Potential reasons why the results in this chapter differ, range from their study relying on variation through time whereas this study heavily relies on cross section variation, over the fact that Carr and Jayadev can identify groups at the state level, to the differences between the PSID and the SCF in their abilities to adequately measure top incomes. Also closely related is the work of Bertrand and Morse (2013). The results of this chapter are consistent with theirs in the sense that both studies report some but limited evidence in favour of the ECH.

Important differences are that Bertrand and Morse (2013) study consumption instead of borrowing behaviour, the fact that they use PSID and CEX data and that they only use the percentile cut-off ($\tilde{Y}_{it}^{pq}$) and do not use average income above the cut-off ($\tilde{Y}_{it}^{avq}$) when defining reference income. Coibion et al. (2016) do not find a positive relationship between income inequality and household borrowing. So while this chapter does not find a dominant role for within group inequality either, the preferred explanations for this lack of a finding which are provided in this chapter and those given by Coibion et al. (2016) are quite different. The latter claim that in general richer households borrow less relative to their income, which is consistent with the results of this chapter. However they also claim that low income households exposed to high inequality borrow less than low income households exposed to a more equal local reference group. They justify this pattern by credit supply shifts and banks relying on income as credit worthiness signals. Altogether the heterogeneity of these studies shows that there is more research to be done. Nevertheless I think the results presented here have the potential to make an important contribution to this debate: Due to its dominant role in explaining household borrowing, the housing market needs to be taken into account in any analysis of household indebtedness.

5 Expenditure Cascades, Income Stagnation or Property Bubbles?

In this chapter the analysis is based on macroeconomic data in contrast to survey data. Macoreconomic data has some important advantages over the SCF: The most important one is that data on household spending as well as information on credit supply conditions is available at the aggregate level. In addition, national accounts and flow of funds data relies to a large extent on government and tax statistics and thus in comparison to survey data is less prone to the problem of misreporting. Although the increasing problem of tax evasion (Zucman 2015) indicates that similar problems might exist on the aggregate level. The most important disadvantage of macroeconomic data is that it is only available at low frequencies, meaning few observations are available to carry out the empirical analysis. This problem is especially relevant when studying the effect of the personal income distribution because these time series are not available at a quarterly frequency[44]. In order to overcome this last complication the approach adopted here is to analyse a panel of similar countries and to pool observations. Overall, using macroeconomic data instead of survey data comes with advantages as well as problems. However in order to obtain a robust test of the various explanations of household debt accumulation, it is necessary to exploit as many data sources as possible. Therefore the analysis presented in the following sections, provides an important and necessary extension of the previous chapters.

The contribution of this chapter is to assess the consistency of the outlined hypotheses with the available macroeconomic data in a unified framework which takes asset prices as well as the distribution of income into account. Prior studies usually focus on one of these two factors. Perugini et al. (2016) for example, do not distinguish between household and business debt and they do not take property prices into account. Bordo and Meissner (2012) and similarly, if critical, Gu and Huang (2014) investigate the effect of inequality on debt over a long period (1920-2008), but do not control for real estate prices or financial regulation. Thus the existing macroeconometric literature lacks a comprehensive empirical study. In order to provide such a study, a household debt equation as well as an auxiliary consumption function are estimated for a panel of 13 OECD countries. The findings can be summarised as follows: There is a very robust direct relationship between residential real estate prices and household debt even after controlling for factors like top income shares and disposable income growth. This is interpreted as support of the HH. In contrast the negative long run impact of income inequality measures on household debt as well as consumption is not in line with the predictions of the ECH nor the ISH. This means that the microeconometric evidence from the previous two chapters is consistent with aggregate data. Once more the main

---

[44] Some authors constructed quarterly measures of the US income distribution (Coibion et al. 2016) but the data sources used are not well suited to identify income inequality at the top of the distribution due to non-observation and differential non-response problems.

conclusion which is to be drawn is that residential real estate plays the dominant role in determining household liabilities.

The rest of the chapter is organized as follows: Section 5.1 distils the key hypotheses. Section 5.2 reviews the relevant empirical literature on the determinants of household debt. Section 5.3 discusses the data sources and the econometric method. Section 5.4 presents the results for the debt equation and section 5.5 for the consumption function. Section 5.6 concludes.

5.1 The Working Hypotheses

In order to assess the consistency of the aggregate data with the three hypothesis of interest, this chapter estimates a debt accumulation and a consumption function. The debt equation treats household borrowing as a function of disposable income (Y^D), property (PP) and stock price (SP) indices, a measure of income inequality (Q), a real interest rate (R), the share of the population which is older than 65 (OLD) and a credit supply index (CS):

$$D = d(Y^D, PP, SP, Q, R, OLD, CS) \tag{7}$$

The consumption function links private consumption expenditures to the same explanatory variables:

$$C = c(Y^D, PP, SP, Q, R, OLD, CS) \tag{8}$$

The debt equation will serve as the primary tool for investigating the hypotheses at stake but the consumption function will provide useful auxiliary evidence. According to the ECH households engage in debt-financed spending in an attempt to emulate the social status of richer peers. In contrast the ISH argues that those households with below average income growth take on debt in order to keep living standards constant. In both cases rising inequality encourages household borrowing: $\frac{D}{\partial Q} > 0$.

The difference between the two is that with the ECH consumption expenditures increase with income inequality. In contrast the ISH is consistent with lower consumption expenditures in periods of high income inequality. The reason is that even if households borrow in order to compensate slower income growth, they are not able to fully compensate sluggish income growth by taking on debt.[45] The argument is that in an environment of above average top income growth, lower income groups decrease their marginal propensities to save and thus the gap in saving propensities along the distribution of income increases. The implicit assumption is that the reduction in saving propensities is enough to outweigh the lower propensity to consume at the top of the income

[45] The expenditure cascade hypothesis is consistent with different propensities to save along the income distribution despite its prediction of a positive consumption effect of an increase in income inequality.

distribution. The ECH implies a positive effect of inequality on consumption ($\frac{\partial C}{\partial Q} > 0$), whereas the income stagnation argument implies a negative effect ($\frac{\partial C}{\partial Q} < 0$).

According to the HH, changes in debt are driven by asset transactions and collateral-backed borrowing due to wealth-effects as well as speculative real estate transactions motivated by the expectation of price increases. Here the key driving variable are property prices: $\frac{\partial D}{\partial PP} > 0$. Because of potential wealth effects as well as liquidity constrained households, the HH also predicts positive effects of property prices on consumption spending: $\frac{\partial C}{\partial PP} > 0$. In addition, financial deregulation has enabled and encourage more lending by financial institutions, leading to more household borrowing ($\frac{\partial D}{\partial CS} > 0$). Table 7 summarises the predictions of the three hypotheses on the debt accumulation and consumption function.

Table 7: Hypotheses on debt determinants

	Hypothesis	Theoretical Argument	Predicted signs
1	expenditure cascades hypothesis (ECH)	Households make consumption decisions with respect to richer peers. Consumption decisions drive debt	$\frac{\partial D}{\partial Q} > 0$ and $\frac{\partial C}{\partial Q} > 0$
2	income stagnation hypothesis (ISH)	Households use debt as a substitute for slow or negative income growth.	$\frac{\partial D}{\partial Q} > 0$ and $\frac{\partial C}{\partial Q} < 0$
3	housing hypothesis (HH)	Debt is driven by asset transactions and wealth effects. Rising asset prices lead to higher debt and higher spending.	$\frac{\partial D}{\partial PP} > 0$ and $\frac{\partial C}{\partial PP} > 0$

D is household debt, Q is a measure of income inequality, C is a measure of aggregate consumption, CS stands for credit supply, R is a real interest rate and PP indicates property prices.

5.2 The Macroeconometric Literature

There are four strands of empirical literature which provide reference points for assessing the three hypotheses outlined in the previous section. The first two deal with the effect of income inequality on household debt and consumption and the other two estimate the impact of changes in housing wealth on household debt and consumption expenditures. Three general patterns emerge when reviewing the literature: First, there is a lack of studies which investigate the impact of income inequality and property prices simultaneously. Starting to fill this gap is one of the main contributions of this paper. Second, the literature very robustly finds a positive impact of real estate prices on household borrowing and consumption which is in line with the HH. Third, the evidence about the impact of the distribution of income on household borrowing and consumption is mixed. No general conclusion can be drawn.

5.2.1 Income Inequality as a Driver of Household Debt and Consumption

The Financial Crisis triggered by the collapse of the US mortgage market unleashed a wave of macroeconometric studies which look at the relationship between the trend of rising income inequality and household indebtedness (Klein 2015; Perugini et al. 2016; Gu & Huang 2014; Malinen 2014; Behringer & Treeck 2013; Bordo & Meissner 2012; Kumhof et al. 2012). Many of these studies are motivated by the theoretical works of Rajan (2010) and Kumhof and Rancière (2010) who argue that rising income inequality contributed to the build-up of a private debt bubble[46]. These papers do not rely on a strict theoretical model but come up with ad hoc specifications which they estimate. All of them rely on aggregate panel data. There is a high degree of heterogeneity with respect to the debt measures used and the specifications of the empirical models and thus a direct comparison to the results presented in this chapter is not always possible.

Perugini et al. (2016), Gu and Huang (2014) and Bordo and Meissner (2012) for example use broad credit measures which also include the liabilities of the corporate sector as their dependent variable. Perugini et al. (2016) use a similar dataset of 18 OECD countries from 1970 to 2007. However they use debt as a share of GDP as their dependent variable and a dynamic system GMM approach with which they find a positive impact of top income shares. Gu and Huang (2014) and Bordo and Meissner (2012) in contrast use long data series going back to the 1920s and use the logarithmized difference of their deflated credit measure as dependent variable. The latter do not find a positive impact of top income shares whereas the latter do claim to find such a relationship but it hinges on interacting the inequality measure which GDP growth. Only Perugini et al. (2016) make an attempt to control for credit supply conditions and find a positive effect of the credit market deregulation index supplied by the Fraser Institute. These studies are different from the analysis carried out in this chapter in three ways. First, none focuses on household debt which is the explicit goal here. Second, none of these studies takes the time series properties of the data fully into account and instead apply microeconometric panel estimation methods. Third, our sample period begins in 1980 instead of 1920 or 1970.

Klein (2015) and Malinen (2014) do not aim to test a specific theoretical model either but are motivated by previous empirical studies and the lack of cointegration tests therein. Both authors investigate bivariate cointegration relationships between household debt (Klein 2015) or bank credit to the private sector (Malinen 2014) and top income shares. They find evidence of a cointegration relationship between debt and income inequality. Klein (2015) estimates the cointegrating vector in a purely bivariate model whereas Malinen (2014) controls for short run fluctuations in GDP, investment and the money stock M2. Both find a positive long run effect of top income shares on private debt. While investigating the long run relationship between household

[46] Several other authors emphasized the role of income inequality in explaining the Financial Crisis even earlier (Palley 2010; Horn et al. 2009; Stockhammer 2009).

borrowing and income inequality is interesting any such relationship most likely will involve more than just two variables. This is why the author refrains from bivariate cointegration tests and instead applies a fully specified dynamic model.

Table 8: Effects of income distribution on household debt

authors	dependent variable	country	findings
Behringer and van Treeck 2013	current account balance in % of GDP, household financial balance in % of GDP, household saving rate	annual data, G7, 1972-2007	top income shares and Gini coefficients have negative effects on dep. vars.
Bordo and Meissner 2012	domestic bank loans to the private sector in real terms*	annual data, 14 OECD countries, 1920-2008	no statistically significant effect of top 1% income shares on dep. var.
Gu and Huang 2014	domestic bank loans to the private sector in real terms*	annual data, 14 OECD countries, 1920-2008	statistically significant positive effect of top 1% income shares on dep. var. if interacted with GDP growth
Klein 2015	credit to private households per capita in real terms	annual data, 9 OECD countries, 1953-2008	statistically significant positive effect of top 1% income share, Gini coefficient, wage share
Kumhof et al. 2012	current account balance in % of GDP	annual data, 18 OECD countries, 1968-2006	statistically significant negative effect of top income shares on current account,
Malinen 2014	domestic bank loans to the private sector in real terms*	annual data, 8 OECD countries, 1960-2008	statistically significant positive effect of top 1% income shares on dep. var.
Perugini et al. 2015	domestic credit to private sector in % of GDP	annual data, 18 OECD countries, 1970-2007	statistically significant positive effects of top 1% / 5% / 10% income shares on dep. var.

* indicates that the paper use the Schularick and Taylor (2012) data set.

Behringer and van Treeck (2013) and Kumhof et al. (2012) do not look directly at private sector or household debt but at the current account balance in % of GDP. The idea is that if households engage in debt financed expenditure cascades due to upward looking status comparison, then the current account will deteriorate given one controls for changes in the corporate and public sector balance. Thus they estimate a model with the current account as dependent variable and top income shares as their preferred measure of income inequality. A negative effect of top income shares on the current account balance is interpreted as an accumulation of liabilities towards the rest of the world and evidence in favour of the ECH. Both studies report negative effects and do not control for credit market supply shifts. Table 8 summarizes the empirical literature investigating the effects of shifts to the distribution of income on household borrowing.

A final but important difference of the analysis of this chapter compared to the works discussed is that it explicitly controls for household assets and especially real estate prices. As the primary residence is the most important assets for the majority of households and the most frequent reason to take on debt, ignoring fluctuations in asset prices raises the potential for serious omitted variable bias problems.

With respect to the impact of the distribution of income on aggregate consumption spending, the macroeconometric evidence is scarce. Most of the more recent studies using aggregate data are motivated by debates in development economics about the relationship between income inequality and saving rates. Alesina and Rodrik (1994) asked whether income inequality might negatively affect saving and growth, which led to a wave of empirical papers (Leigh & Posso 2009; Li & Zou 2004; Smith 2001; Schmidt-Hebbel & Servén 2000). All of them use aggregate panel data and investigate the impact of income inequality on aggregate saving and report results for OECD countries[47]. Due to the inverse relationship between saving and consumption these findings are also relevant for investigating the ECH and the ISH. The overall picture is inconclusive: Leigh and Posso (2009) and Schmidt-Hebbel and Servén (2000) do not find any consistent relationship between the two variables. Li and Zou (2004) on the other hand find negative effects of income inequality on the aggregate saving rate and Smith (2001) finds positive effects. There are also a few recent papers studying the effects of income inequality on saving rates (Alvarez-Cuadrado & Vilalta 2012) and consumption (Stockhammer & Wildauer 2016) which are explicitly motivated by the ECH. Alvarez-Cuadrado and Vilalta (2012) find negative effects of income inequality on saving rates in an OECD panel while Stockhammer and Wildauer (2016) report positive but very small and statistically insignificant effects of top income shares on consumption expenditures. Once again real estate assets are omitted from all of these studies, except in Stockhammer and Wildauer (2016).

5.2.2 The effect of property prices on household debt and consumption

Most of the macroeconometric literature investigating the linkages between property prices and household sector borrowing relies on time series methods. The higher frequencies at which data on house prices is available compared to information about the distribution of income allows researchers to adopt techniques which explicitly model the interdependencies and feedbacks

[47] The author does not consider papers which are based on samples also including developing countries due to the limited comparability to the sample used in this chapter. The author would also be concerned about the validity of the panel homogeneity assumption in highly heterogeneous panels which include developed as well as developing countries.

between the variables involved and allow for cointegration relationships. Thus the standard framework used to investigate the determinants of household borrowing is that of cointegrated vector autoregressions (CVAR) (Hofmann 2004; Chrystal & Mizen 2005; Oikarinen 2009; Gimeno & Martinez-Carrascal 2010; Anundsen & Jansen 2013; Meng et al. 2013). These authors use data on 16 OECD countries, the UK, Finland, Spain, Norway and Australia, respectively. All of them report a statistically significant and positive long run impact of property prices or some form of household wealth on household borrowing, as theory predicts. Hofman (2004) is the only study which uses bank credit to the private sector instead of a narrow definition of household borrowing. Christen and Morgan (2005) and Arestis and Gonzales (2014) rely on single equation cointegration analysis for the US case and a sample of 9 OECD countries, respectively. Also the single equation studies find statistically significant positive long run effects of property prices on household borrowing.

None of these studies makes an effort to control for shifts in credit supply conditions. The only exception is Oikarinen (2009) who splits the Finnish sample into pre and post-1989 periods and argues that doing so allows to control for major changes in the regulatory framework of the financial sector. The findings indicate that only in the period of financial liberalisation property prices affect household borrowing. The reluctance of the literature to take credit supply serious is rooted in the fundamental problem that it is not straightforward how to adequately measure such a broad concept as 'credit supply'. Focussing on the regulatory framework and using proxies like the proportion of publicly owned banks or interest rate spreads as a measure of competition do not necessarily translate into a measure of the willingness to lend. On the other hand if one uses loan-to-value and loan-to-income ratios in order to construct credit conditions indices as Fernande-Corugedo and Muellbauer (2006) do, one adds little information if borrowing and income are already part of the VAR.

The literature using panel data sets is more diverse in the methods applied. Since these studies are also directly comparable to the analysis of this chapter, they are reviewed in more detail. Égert et al. (2006) estimate the determinants of credit to the private sector over GDP in order to assess whether debt levels in central and eastern European countries are in line with long run equilibrium estimates. They use simple fixed effects models as well as the mean group estimator (Pesaran et al. 1999) and dynamic OLS and find a significant and positive effect of house prices on private credit. They also include the spread between lending and deposit rates as a proxy for competition within the banking sector but it remains statistically insignificant as long as house prices are included in the regression. Goodhart and Hofmann (2008) estimate a panel VAR based on a sample of quarterly data from 1970-2006 of 17 OECD countries. The VAR includes nominal bank credit to the private sector, nominal house prices, real GDP, the CPI, nominal interest rates and the money aggregate M3. Based on Granger causality testing and a simple Cholesky decomposition (ordering: GDP, CPI, interest rate, property prices, money, private credit) they find multidirectional links between these variables. They argue in particular that house prices positively influence private credit and money. In order to account for shifts in the regulatory framework of credit markets they re-estimate the model on the shorter period 1986-2006 and find particularly strong effects of house prices. Rubaszek and Serwa (2014) build a theoretical model of household borrowing and compare it with single equation cointegration estimations based on a panel of 36 countries. They use household borrowing relative to GDP as their explanatory variable and find a positive long run impact of house prices on household borrowing. The spread between lending and deposit rates is used as a measure of banking sector competition. This chapter differs from these studies in two important aspects: First, it explicitly uses a measure of household sector borrowing in contrast to credit to the private sector. This is important because the channels influencing household and corporate borrowing are quite different and this thesis is only interested in the former group. Second, none of the above studies takes the distribution of income into account. However, if the ECH or the ISH are valid, ignoring income inequality will lead to biased estimates. Furthermore taking the distribution of income into account allows the author to test various hypotheses in a single regression framework. Table 9 summarizes the literature investigating the nexus between real estate prices and household borrowing.

Attanasio and Weber (1994) argue that (in the UK) the link between house prices and consumption is due to shifts in households expectations about future income (Attanasio et al. 2009) and that aggregate data cannot take this latter explanation properly into account. Instead they interpret the fact that young households react stronger to regional house price increases than older households as evidence for shifts in expected productivity. The alternative explanation of this result that younger households are more likely to be credit constrained, is denied on the grounds that interacting house prices with an ownership dummy yields a similar coefficient for young and old households while the large and positive effect of the ownership dummy is ignored. The Financial Crisis delivered the ultimate blow to this at best bold interpretation of the empirical evidence: If productivity gains had been the driving force of house prices and consumption they would not have collapsed so strongly. Thus we do not pay further attention to this line of reasoning[48].

[48] Also see Muellbauer and Murphy (2008) and Aron et al. (2012) for a critique of the common factor view.

Table 9: Effects of property prices on household debt

Authors	model/debt measure	Country	findings
Time series			
Anundsen and Jansen 2013	structural VEC model; real household liabilities	Norwegian data, 1986Q2 to 2008Q4	positive effect of real house prices on household debt
Arestis and Gonzalez 2014	Error Correction models; domestic bank credit to the private sector as % of GDP	9 OECD countries, 1970-2011	positive long run impact of real property prices on bank credit
Chrystal and Mizen 2005	CVAR including unsecured household sector credit (excl. mortgages)	UK data, 1979Q1 to 1998Q4	positive long run impact of household real net total wealth on household credit
Gimeno and Martinze-Carrasca 2010	CVAR including real house purchase loans per household	Spanish data, 1984Q1 to 2009Q1	house purchase loans depend positively on real house prices
Hofmann 2004	CVAR including domestic credit to the private sector	16 OECD countries, 1980-1998	in almost all countries cointegration relationship with positive effect of real property prices
Meng et al. 2013	CVAR including nominal household liabilities	Australian data, 1988Q2 to 2011Q2	positive effects of nominal house prices on nominal household borrowing
Oikarinen 2009	CVAR including household liabilities as % of GDP	Finish data, quarterly 1975-2006	positive impact of real house prices on household borrowing only after financial liberalization in 1989
Panel data			
Egert et al. 2006	domestic credit to private sector as % of GDP	annual panel data 43 countries, 1975-2004	find a significant and positive effect of nominal house prices on private credit
Goodhart and Hofmann 2008	panel VAR including nominal bank credit to the private sector	17 OECD countries, 1970-2006, quarterly data	interdependency between nominal house prices and household borrowing
Rubaszek and Serwa 2014	household liabilities as % of GDP	36 countries, 1995-2009	positive and highly statistical significant effect of real property prices on borrowing

5.3 Data and Econometric Method

The dataset used is an unbalanced annual panel covering 13 countries from 1980 to 2011[49]. Definitions and data sources are provided in Appendix IV in Table 1 and descriptive statistics in Table 2. While most of the data series are fairly standard national accounts concepts, three kinds

[49] The countries included are: Australia, Belgium, Canada, Germany, Finland, France, Italy, Japan, Netherlands, Norway, Sweden, United Kingdom and the US.

of variables deserve further comment: the asset price indices, the measure of income inequality and the credit supply indices.

First, real residential property and stock price indices are used as proxies for housing and financial wealth of the household sector, because wealth data are not available (for sufficiently long time periods) for most countries. This is common in the literature estimating wealth effects[50], but it only captures price changes and does not capture quantity changes. Second, two different measures of the distribution of income are used. The share of total income received by the richest 1% of households ($TOP1$) by definition mainly captures the dynamics at the top of the distribution. Since the ECH predicts debt-financed spending sprees to be triggered by concentration of income at the top this is the preferred measure for testing that hypothesis. In addition a Gini coefficient which is directly computed from income data ($GINI$) is used. The Gini index is expected to be less sensitive to distributional changes at the top. Third, two different credit supply measures enter the regressions. Although shifts in credit supply conditions are important determinants of household borrowing, measuring the state of credit supply and the willingness to lend by financial institutions is tricky. In the literature one of two approaches is adopted. The first focuses entirely on credit regulations and financial reforms and argues that a less regulated financial sector should be expected to enhance borrowing. Indices based on the existence of interest rate controls, the relation of public to private borrowing, entry barriers to the financial sector and the existence of capital account restrictions are derived and used in empirical analysis. Two very common indices following such an approach are the Fraser Index on credit regulation and the financial reform index by Abiad et al. (2008). The second approach aims to capture shifts in banks' willingness to lend not simply by taking the regulatory framework into account but by creating a direct measure of the risk appetite of the financial sector. Fernandez-Corugedo and Muellbauer (2006) for example back out a common trend from demand for mortgages, unsecured debt and from the evolution of the fraction of high loan-to-value and loan-to-income borrowers in the UK. This chapter uses the Fraser Index as well as the financial reform index of Abiad et al. (2008). Carefully constructed credit supply indices in the spirit of Fernandez-Corugedo and Muellbauer (2006) are not available in a consistent form for several countries. The problem with the indices provided by the Fraser Institute and by Abiad et al. (2008) is that they do not capture shifts in the risk appetite of the financial sector. They also do not capture those changes to the regulatory framework which turned out to be key for the pre-crisis period: the use of off-balance sheet vehicles, increased proprietary trading and low capital requirements for assets in the trading book. For these reasons the author uses both indices in the empirical analysis to achieve results which are comparable to the literature but is rather pessimistic about their ability to effectively measure shifts in credit supply conditions.

[50] See Paiella (2009), Attanasio and Weber (2010) and Cooper and Dynan (2016) for recent surveys.

For estimating the behavioural equations introduced in section 5.1 two different models are used. The first type is sometimes referred to in the literature as growth regressions or differenced data models (Hendry 1995, p.232). This means the main series (disposable income, household debt and consumption) are differenced prior to estimation because they contain unit roots[51]. Due to logarithmic transformation of the data, the model can then be interpreted as regressing growth rates on each other. A simple bivariate example would be:

$$\Delta y_{it} = \alpha_i + \beta_2 \Delta x_{it} + \mu_{it} \tag{9}$$

where i indicate countries, t time and μ_{it} is an error term and the model is estimated with fixed effect (within estimator). This approach is completely unrelated to the empirical literature on the Solow growth model which is also referred to as growth regressions. Following Hendry (1995, p.232) one can think of equation (9) as a restricted version of the autoregressive distributed lag model:

$$y_{it} = \alpha_i + \beta_1 y_{it-1} + \beta_2 x_{it} + \beta_3 x_{it-1} + \mu_{it} \tag{10}$$

if $\beta_1 = 1$ and $\beta_2 = -\beta_3$. Relaxing these implicit restrictions allows for a richer set of dynamics but also requires twice as many parameters to be estimated. Because distributional data is only available at annual frequencies and for a limited number of countries, the restricted model will serve as the pragmatic starting point for the empirical analysis. In addition to fixed effects estimation of equation (9), the mean group (MG) estimator developed by Pesaran and Smith (1995) is applied to equation (9) to obtain a simple check of the assumption of homogeneous effects of the regressors across countries. With the MG estimator equation (9) is estimated separately for each country by OLS. Then point estimates are averaged and reported as MG estimates. Pesaran and Smith (1995, p.96) do not derive precise standard errors for their estimates but instead assume that the regression coefficients are independently distributed across groups (countries in our application). In practice this means regressing the group specific coefficients on a constant which yields the mean and under the assumption of independently distributed coefficients also an estimate of the standard error which the authors then use for hypothesis testing. The Stata command used to carry out the MG regressions was "xtmg" which is described in Eberhardt (2012).

The second model relies on a fully dynamic structure as in equation (10) which can be rewritten in error correction form as:

$$\Delta y_{it} = \alpha_i - \beta_3 \Delta x_{it} - (1 - \beta_1)\left(y_{it-1} - \frac{\beta_3 + \beta_2}{1 - \beta_1} x_{it}\right) + \mu_{it} \tag{11}$$

[51] The author performed first generation unit root tests following Choi (2001). Results are presented in Appendix IV in Tables 3 and 4.

It is important to note that including the contemporaneous value x_{it} in the long run relation[52] might be confusing to some authors. In fact this is just one way to reparametrize equation (10). Most importantly reparametrizing it such that x_{it-1} appears in the long run trend does not change the long run coefficients and their interpretation. However it would not allow for the situation with only one lag of the dependent variable and no other lag as a special case. This is the reason why the author followed the literature (Blackburne & Frank 2007; Pesaran et al. 1999) and used the reparametrization with x_{it} instead of x_{it-1} in the long run part of the model. As a compromise between allowing for cross country heterogeneity and restricting the number of parameters which need to be estimated, the long run coefficients of the corresponding error correction model are restricted to be homogeneous across countries. This is the Pooled Mean Group (PMG) model of Pesaran et al. (1999). Allowing the short term parameters β_3 and the adjustment speed $(1-\beta_1)$ to be heterogeneous across countries, while the long run equilibrium relationship $\frac{\beta_3+\beta_2}{1-\beta_1}$ is homogeneous across countries yields the PMG model of equation (11):

$$\Delta y_{it} = \alpha_i - \beta_{3i}\Delta x_{it} - \phi_i(y_{it-1} - \lambda x_{it}) + \mu_{it} \tag{12}$$

However since the long run parameters are a non-linear function of the short run parameters the PMG imposes the following restrictions on the model:

$$\beta_2 = \lambda\phi_i - \beta_{3i} \tag{13}$$

Pesaran et al. (1999) argue that based on theoretical grounds the long run income elasticity of consumption should be one across all countries because otherwise saving rates would rise or fall indefinitely. According to them imposing common long run restrictions in situations where economic theory suggests them allows to reduce the necessary parameters to be estimated significantly. For example, estimating an unrestricted autoregressive distributed lag model $ARDL(p_0, p_1, \dots, p_k)$ with common lag length p, k regressors and N countries, requires to estimate $2N + (p+1)kN$ parameters. In contrast estimating the PMG version of that model only requires $2N + k(pN+1)$ parameters. With 13 countries, 5 regressors and 1 lag this amounts to 156 and 96 parameters, respectively. Another advantage of the PMG estimator is that it allows for I(0) as well as I(1) variables and requires no prior knowledge of the order of integration. In order to determine the lag structure of the error correction model the author applies a testing down procedure. First a fully specified model including one lag of all independent variables and two lags of the dependent variable is estimated and then statistically insignificant short run effects are removed.

[52] We prefer to call the cointegrating vector a long run trend rather than a long run equilibrium. The reason is that our findings indicate that the debt accumulation trend prior to 2007 was unsustainable, making the notion of an equilibrium ill-suited in order to refer to such a trend.

Pesaran et al. (1999, p.624) derive a maximum likelihood estimator for equation (13) which first estimates the homogeneous long run relationship (λ) and the country specific adjustment speeds (ϕ_i) by maximizing the likelihood function of the model which is the product of the individual country likelihood functions. After iteratively estimating $\{\phi_i, \lambda\}$, the short run coefficient(s) (β_{3i}) are estimated by running country specific OLS regressions of the form (Pesaran et al. 1999, p.626):

$$\Delta y_{it} = \alpha_i - \beta_{3i}\Delta x_{it} - \phi_i\hat{\xi} + \mu_{it} \tag{12b}$$

where $\hat{\xi} = y_{it-1} - \hat{\lambda}x_{it}$. The Stata routine used to implement this estimation procedure is "xtpmg" which is described in Blackburne and Frank (2007).

The preferred sample covers the period 1980-2007 and excludes observations from the post-crisis period. In principle one can argue that the effects associated with the different hypothesis which the thesis aims to test, should be symmetric in the sense that they apply in the upswing as well as in the downswing of the business cycle. However there are good reasons why this symmetry will not hold in practice. Most importantly, it is easier for households to accumulated debt and thus will happen quicker, rather than it is to deleverage. Especially to deleverage during an economic downturn requires more time than leveraging up in the upswing. Formally such arguments should be checked by testing whether the financial crisis marks a structural break point in the time series. Due to the limited amount of post crisis data available this is difficult to do. However the vast literature which tries to explain why growth remains low and sluggish even years after the crisis (Barkbu et al. 2015; European Commission 2015; Lewis et al. 2014) provides evidence for such a break or shift in the long term trend (and thus in the cointegrating vector). Assessing these questions in more detail is an interesting task for future research.

5.4 Determinants of Household Debt

The first step is to estimate household liabilities (D_{it}) as a function of real disposable household income (Y_{it}^D), the income share of the richest 1% of households ($TOP1_{it}$), alternatively a Gini coefficient ($GINI_{it}$) of income inequality is used. Furthermore real property prices (PP_{it}) and real stock prices (SP_{it}), the real long term interest rate (R_{it}), the ratio of people older than 65 in the population (OLD_{it}) and either the financial reform index from Abiad et al. (2008) (FIN_{it}) or the credit market regulation index published by the Fraser Institute ($CRED_{it}$) (subcategory 5A of the Economic Freedom of the World index) enter the regression. Due to the non-stationarity of these series, all series are differenced and a fixed effects estimator is applied. The following equation is estimated, lower case letters indicate the underlying variables are transformed by taking natural logarithms:

$$\Delta d_{it} = \alpha_i + \beta_1\Delta y_{it}^D + \beta_2\Delta Q_{it} + \beta_3\Delta pp_{it} + \beta_4\Delta sp_{it} + \\ \beta_5\Delta R_{it} + \beta_6\Delta OLD_{it} + \beta_7\Delta s_{it} + \mu_{it} \tag{14}$$

where Q stands for one of two measures of income inequality and s represents one of the two credit supply indices. Equation (14) essentially relates the growth rate of household debt to the growth rate of disposable income, property and share prices and changes in the level of income inequality, the real interest rate, demography and the stance of credit supply. By allowing heterogeneous intercepts in equation (14) we allow for heterogeneous time trends in the level of household debt across countries[53].

Table 10 reports the results from estimating equation (14). In columns (1) to (4) results based on the fixed effects (FE) estimator are reported. The coefficient on disposable income is statistically significant at the 1% level, with a magnitude around 0.5 indicating that households take on more debt as incomes and the economy grow. Next, there is an equally statistically significant positive coefficient on property prices indicating that a 1% increase in real estate prices increases household borrowing by 0.36% on average in this sample of 13 OECD countries. The positive coefficient on stock prices is much smaller (around 0.02) compared to the property price coefficient and not statistically significant. Finding positive and statistically significant coefficients on asset price indices supports the HH: Asset price growth and in particular real estate price growth are important drivers of household debt. The coefficient on the distribution of income depends on the measure used. There is a statistically significant negative coefficient on the income share of the richest 1% of households which is in stark contrast to the positive effect predicted by the ECH as well as the ISH. Since debt-financed spending cascades should start at the top of the income distribution the top income share should pick up such a mechanism. In contrast, the Gini coefficient which is not as sensitive to changes at the top of the distribution, does not exhibit statistically significant coefficients. These results support the HH but are inconsistent with the ECH as well as the ISH.

Neither of the two credit supply indices exhibits statistically significant coefficients. However the highly significant intercept indicates that household debt exhibits a positive deterministic time trend which is not explained by the included variables. This residual growth of household debt amounts to 2.3% per year. Since neither the IMF’s nor the Fraser Institute’s credit supply index is providing an adequate measure of the willingness to lend or the risk appetite of the financial sector, any such long term shifts in financial sector behaviour are potential explanations for this unexplained time trend.

The coefficient on the real interest rate is not statistically significant. The positive sign is not expected. Higher borrowing costs should have a dampening effect on households’ credit demand. Only in column (4) there is a statistically significant negative coefficient on the old-age ratio. The

[53] A constant component of a growth rate can be interpreted as a time trend similar to a random walk with drift which exhibits a deterministic time trend due to the drift term.

negative sign is in line with economic theory given that major outlays such as buying a home or having children occur before 65 years of age for most households.

As a robustness check equation (14) was estimated separately for each country by OLS and then the results were averaged. These results, referred to as mean group (MG) estimations, are reported in Table 10 in columns (5) and (6). The results for each individual country are reported in Table 5 in Appendix IV. The MG estimator confirms the results from the fixed effects estimation except that it does not yield a statistically significant negative coefficient on the top 1% income share. Taking a look at the individual country results reveals that there are statistically negative coefficients only in Australia and Norway whereas in Italy it is statistically significant and positive.

An additional robustness check is presented in column (7) of Table 10. There the sample is split based on the degree of property price increases as it might be argued that inequality induced borrowing might depend on real estate price increases due to credit constraints. The results of column (7) are solely based on countries which experienced increases in real residential property prices of more than 80% over the sample period (AU, BE, CA, FI, FR, IT, NL, NO, SE, UK, US). Germany and Japan are the only countries in the sample which did not experience pronounced house price booms over the sample period. The results are very close to the full sample results since only Germany and Japan were excluded. This split reveals that even in those countries with especially high increases in house prices, there is no positive direct link between income inequality and household borrowing.

The last robustness check addresses the potential two-way causation between household debt growth and residential property price growth. It is not only plausible to think of real estate as collateral for household borrowing. One can also argue that an increase in household borrowing can boost demand for residential real estate and thus house prices. In order to obtain a quick check to what extent this second explanation is driving the results, equation (14) is re-estimated with residential property prices entering the equation with a lag. Results are reported in Table 6 in Appendix IV. The coefficient on lagged house price growth is about the same magnitude compared to the baseline specification in Table 10 and statistically significant at the 1% level. The remaining results are qualitatively in line with the previous findings. Thus the findings reported in Table 10 are not driven by reverse causality between household debt and residential property prices.

Table 10: Debt growth regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	FE, all countries				MG, all countries		FE, PP↑
Δy^D	0.49***	0.46***	0.48***	0.40***	0.46***	0.46***	0.44***
	(0.12)	(0.12)	(0.11)	(0.12)	(0.16)	(0.13)	(0.12)
Δpp	0.36***	0.36***	0.36***	0.38***	0.38***	0.36***	0.36***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.08)	(0.08)	(0.03)
Δsp	0.02	0.02	0.02	0.02	0.02	0.01	0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
$\Delta TOP1$	-0.47**		-0.61***		1.67		-0.58***
	(0.20)		(0.12)		(1.79)		(0.13)
$\Delta GINI$		0.13		0.11		0.08	
		(0.19)		(0.17)		(0.17)	
ΔR	0.07	0.11	0.06	0.12	-0.15	-0.18	0.05
	(0.06)	(0.07)	(0.06)	(0.07)	(0.09)	(0.13)	(0.06)
ΔOLD	-1.99	-2.04	-2.34	-2.44*	2.95	2.55	-1.10
	(1.67)	(1.58)	(1.39)	(1.34)	(5.58)	(4.94)	(2.60)
Δfin	0.01	0.00					
	(0.04)	(0.04)					
$\Delta cred$			-0.02	-0.02	-0.06	-0.05	-0.02
			(0.04)	(0.04)	(0.10)	(0.09)	(0.05)
$const$	0.03***	0.03***	0.03***	0.03***	0.03***	0.03***	0.03***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)
N	293	293	319	319	320	320	269
adj. R^2	0.41	0.41	0.43	0.42			0.40

Fixed effects (FE) and mean group (MG) estimators, dependent variable: Δd_{it}, robust standard errors clustered at the country level in brackets. Stars indicate statistical significance: * p<0.1, ** p<0.05, *** p<0.01. Lower case letters indicate natural logarithms. Specifications (1) and (2) are based on data from 1980-2005, all other specifications based on data from 1980-2007. Specifications (1) to (6) include all 13 countries of the sample (AU, BE, CA, DE, FI, FR, IT, JP, NL, NO, SE, UK, US). Specification (7) includes countries which experienced increases of real residential property prices by more than 80% over the sample period (AU, BE, CA, FI, FR, IT, NL, NO, UK, US, SE).

The second part of the analysis relaxes some of the dynamic restrictions imposed by equation (14) and estimates an error correction model with heterogeneous short-term effects as well as heterogeneous adjustment speeds. In order to reduce the number of parameters to be estimated, stock price indices are removed from the specification. The rationale for excluding stock prices is that for most households real estate and not equities represent the main form of wealth. Also by far the biggest component of household debt is mortgage debt, which further points to the relatively more important role of real estate compared to financial assets. This latter point is also supported by the statistically insignificant coefficients in the growth regressions. This leaves six explanatory variables. The PMG approach then requires the estimation of 123 parameters. The resulting model is the following:

$$\Delta d_{it} = \alpha_i + \beta_i^1 \Delta y_{it}^D + \beta_i^2 \Delta Q_{it} + \beta_i^3 \Delta pp_{it} + \beta_i^4 \Delta R_{it} + \beta_i^5 \Delta OLD_{it} + \beta_i^6 \Delta s_{it} + \beta_i^7 \Delta d_{it-1}$$

$$-\gamma_i \left(d_{it-1} - \beta^8 y_{it}^D - \beta^9 Q_{it} - \beta^{10} pp_{it} - \beta^{11} R_{it} - \beta^{12} OLD_{it} - \beta^{13} s_{it} \right) + \mu_{it} \qquad (15)$$

Table 11 presents the results of estimating equation (15). In columns (1) and (2) the sample is restricted to the pre-crisis period 1980-2007. The first specification relies on the top 1% income share as an indicator for the degree of income concentration while specification (2) relies on a Gini coefficient. In both cases there is a statistically significant adjustment towards the estimated long run trend. This long run trend is characterized by an income elasticity larger than 1 which implies that even with all other variables constant, debt to income ratios increase as the economy and thus disposable household income grows. The hypothesis that the income coefficient is equal to 1 can only be rejected at a 10% significance level. Since debt cannot grow faster than disposable income forever because debt service payments need to be paid out of household income, this result indicates the unstable or bubble character of this estimated long run trend. With respect to residential property prices the long run elasticity of debt is around 0.56 in both cases and highly statistically significant. Thus the estimated long run real estate elasticity is consistent with the HH. Both coefficients on the income distribution variables are negative and statistically significant at the 1% level. This indicates, when controlling for disposable income, residential real estate prices, long term real interest rates, demography and credit supply conditions, a more polarised distribution of income leads to less household borrowing. This finding contradicts the ECH as well as the ISH which predict a positive relationship. The long run real interest rate elasticity is negative and statistically significant at the 1% level. This finding is in general in line with economic theory, as lower real interest rates are expected to boost household borrowing. The impact of the old-rage ratio is statistically significant and negative in the first specification, which is in line with basic life-cycle interpretations but is also consistent with the interpretation that older households have fewer resources and participate to a lesser extent in economic life. Finally the long run coefficient on the credit market regulation index is positive and highly statistically significant. This latter finding is in line with the notion that household debt expanded due to shifts in credit market regulations and the increased willingness to lend by the financial sector.

Table 11: Household debt, error correction models, pooled mean group (PMG) estimator

	(1)	(2)	(3)	(4)	(5)	(6)
sample	1980-2007		1980-2011		1980-2007, excl. DE & JP	
y^D_{t-1}	1.185***	1.248***	0.781***	0.989***	1.170***	1.223***
	(0.11)	(0.14)	(0.11)	(0.29)	(0.11)	(0.14)
pp_t	0.561***	0.579***	0.636***	-0.053	0.567***	0.589***
	(0.07)	(0.08)	(0.06)	(0.24)	(0.07)	(0.08)
$TOP1_t$	-1.390***		0.274		-1.361***	
	(0.46)		(0.72)		(0.46)	
$GINI_t$		-3.344***		-5.718***		-3.221***
		(0.78)		(2.21)		(0.77)
R_t	-0.845***	-1.732***	-0.928**	-5.249***	-0.854***	-1.710***
	(0.22)	(0.29)	(0.40)	(1.21)	(0.22)	(0.28)
OLD_t	-3.435***	1.618	1.446*	16.653***	-3.485***	1.472
	(0.90)	(1.05)	(0.86)	(3.84)	(0.90)	(1.04)
$cred_t$	0.335***	1.310***	0.828***	5.033***	0.346***	1.296***
	(0.12)	(0.18)	(0.13)	(1.14)	(0.12)	(0.17)
	short run		short run		short run	
$adjustment$	-0.137***	-0.072***	-0.104***	-0.014	-0.155***	-0.089***
	(0.04)	(0.02)	(0.02)	(0.01)	(0.04)	(0.02)
Δy^D_t	0.131	0.230**	0.162	0.279***	0.078	0.210*
	(0.11)	(0.10)	(0.11)	(0.10)	(0.12)	(0.12)
Δpp_t	0.138**	0.166**	0.170***	0.230***	0.161**	0.189***
	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)	(0.06)
$\Delta TOP1_t$	1.1		0.317		1.011	
	(0.98)		(0.91)		(1.08)	
$\Delta GINI_t$		0.04		0.027		0.082
		(0.21)		(0.21)		(0.24)
ΔR_t	0.016	0.075	0.131**	0.07	-0.011	0.061
	(0.07)	(0.06)	(0.05)	(0.08)	(0.08)	(0.07)
ΔOLD_t	7.786	4.012	4.642	0.025	9.422*	5.551
	(4.89)	(3.22)	(3.45)	(1.50)	(5.72)	(3.69)
$\Delta cred_t$	-0.169***	-0.149***	-0.194***	-0.113***	-0.192***	-0.174***
	(0.05)	(0.03)	(0.04)	(0.04)	(0.06)	(0.03)
Δdh_{t-1}	0.446***	0.445***	0.511***	0.418***	0.430***	0.441***
	(0.07)	(0.08)	(0.06)	(0.06)	(0.07)	(0.09)
$constant$	-0.168***	-0.216***	-0.047**	-0.133	-0.176***	-0.257***
	(0.06)	(0.07)	(0.02)	(0.12)	(0.06)	(0.08)
N	306	306	357	366	259	259
$H_0: \beta_{y^D} = 1$	0.09	0.07	0.05	0.97	0.12	0.10
$H_0: \beta_{PP} = 1$	0.00	0.00	0.00	0.00	0.00	0.00

Error correction models with homogeneous long run effects and heterogeneous short run effects and adjustment speeds. Dependent variable: Δd_{it}. Stars indicate statistical significance: * p<0.1, ** p<0.05, *** p<0.01. Lower case letters indicate variables are transformed by taking natural logarithms. Standard errors in brackets.

When the sample period is extended to also include the post-crisis years (1980-2011) as is the case in columns (3) and (4), there is no statistically significant adjustment towards the estimated long run trend when the Gini coefficient is used. When the top 1% income share is used, there is a statistically significant adjustment but the long run trend towards which this adjustment occurs is different, compared to the short sample. Most importantly the long run income elasticity is smaller than 1, the long run elasticity of the top income share is not statistically significant and the long run old-age coefficient is positive. Given that the aim of the chapter is to assess the drivers of household debt in the pre-crisis period these two specifications are interpreted as robustness checks. A likely explanation for the changed results with the extended sample is that the economic crisis and its aftermath exhibit a fundamentally different trend compared to the pre-crisis period. Theoretically this means there are hysteresis effects at work. Indirect empirical support for this assessment is provided by the many empirical studies which investigate why investment (and growth) did not pick up after the crisis (Barkbu et al. 2015; European Commission 2015; Lewis et al. 2014).

In columns (5) and (6) the sample only includes observations up to 2007 but excludes Germany and Japan. The reason for excluding these two countries is that they experienced very different real estate market dynamics compared to the rest of the sample: While in most countries house prices picked up in the 80s or 90s and peaked around 2007, in Germany they remained flat throughout the entire period while they peaked in 1991 in Japan. Therefore, excluding these two countries enables the author to identify differences in the long run trend between countries which experienced rapidly growing house prices in the decade prior to the financial crisis and those which did not. Because only two countries are removed from the sample, results in columns (5) and (6) are almost identical to the results in columns (1) and (2). There is one important difference however: the long run income elasticity is statistically different from 1 with coefficients of 1.17 and 1.22, respectively. This latter result emphasises the unsustainable nature of household debt accumulation in countries with rapidly rising real estate prices.

In addition to the statistical significance of the estimated coefficients, also the economic significance is relevant in assessing the explanatory power of the three hypotheses. Economic significance was computed first for the entire panel based on the estimated long run coefficients. In order to assess the contribution to the change in household debt between 1995 and 2007 of each regressor the following computation was made after GDP weighted cross section averages were taken and household debt and all series were transformed into chained purchasing power parity 2005 Dollars:

$$\begin{aligned}\ln\left(\frac{\widehat{D}_{2007}}{\widehat{D}_{1995}}\right) &= \hat{\beta}^{8}\ln\left(\frac{Y^{D}_{2007}}{Y^{D}_{1995}}\right) + \hat{\beta}^{9}(Q_{2007} - Q_{1995}) + \hat{\beta}^{10}\ln\left(\frac{PP_{2007}}{PP_{1995}}\right) \\ &+ \hat{\beta}^{11}(R_{2007} - R_{1995}) + \hat{\beta}^{12}(OLD_{2007} - OLD_{1995}) + \hat{\beta}^{13}\ln\left(\frac{S_{2007}}{S_{1995}}\right)\end{aligned} \tag{16}$$

After some manipulation equation (16) becomes:

$$\frac{\widehat{D}_{2007}}{\widehat{D}_{1995}} = \left(\frac{Y^D_{2007}}{Y^D_{1995}}\right)^{\widehat{\beta}^8} e^{\widehat{\beta}^9(Q_{2007}-Q_{1995})} \left(\frac{PP_{2007}}{PP_{1995}}\right)^{\widehat{\beta}^{10}} e^{\widehat{\beta}^{11}(R_{2007}-R_{1995})} e^{\widehat{\beta}^{12}(OLD_{2007}-OLD_{1995})} \left(\frac{S_{2007}}{S_{1995}}\right)^{\widehat{\beta}^{13}} \quad (17)$$

From equation (17) the contributions for each individual variable can be defined. For example in the case of disposable household income:

$$\frac{\widehat{D}^y_{2007}}{\widehat{D}^y_{1995}} = \left(\frac{Y^D_{2007}}{Y^D_{1995}}\right)^{\widehat{\beta}^8} \quad (18)$$

Equation (18) represents the change in household debt between 2007 and 1995 which can be attributed to the growth of disposable income according to the average country in our model. The contributions to household debt between 2007 and 1995 for all variables are presented in Table 12 which also contain separate results when making these calculations not with averaged data but with US data.

Table 12: Contributions to changes in household debt between 1995 and 2007

	actual change in D	predicted change in D	Y^D	$TOP1$	PP	R	OLD	$CRED$
panel	145%	126%	64%	-5%	43%	3%	-2%	1%
US	126%	116%	61%	-6%	42%	2%	0%	-1%

The predicted change in debt is computed based on equation (17). Contributions of individual variables are computed based on equation (18). Calculations used the estimated coefficients from column 5 in Table 11. Results for the panel were obtained by taking GDP weighted averages across countries. The product of the individual change factors (i.e. 1.6 instead of 60%) yields the predicted change in D.

Table 12 provides three clear results: First, debt grows as disposable income and the economy grow. This is not a particularly surprising result. It is important to note however that the estimated model predicts debt growth in excess of income growth even after controlling for other factors. This means that households became increasingly indebted relative to disposable income due to factors which are not part of the model. Shifts in attitudes towards borrowing could be an explanation as well as the inadequacy of the credit supply measures to represent changes in banking practices and increased risk seeking by the financial sector. Second, real appreciations of residential property prices explain almost the entire change in household debt after controlling for disposable income. This means that the main driver of debt to income ratios were strongly increasing real estate prices. Third, real interest rates, demographic shifts and changes in credit market regulation played a negligible role for household borrowing outcomes according to the estimated model.

Based on the growth regressions and Error Correction Models, three conclusions can be drawn: First, the larger than unity long run income elasticities indicate that debt accumulation was unsustainable in the core sample, excluding Germany and Japan. The finding of positive income elasticities is an expected result as households borrow more as their disposable income and the economy grow. However larger than unity long run elasticities indicate growing debt to income ratios even when holding other important drivers of debt such as property prices constant. Such a long run trend cannot be sustained indefinitely and thus demonstrates the bubble character of household debt accumulation prior to the 2007/2008 crisis. Second, property prices exhibit substantial explanatory power for predicting growing household debt . This relationship likely reflects three mechanisms at work. On the one hand households borrow more relative to their income in order to afford rising house prices. On the other hand owners will re-mortgage or take out additional mortgages in order to 'cash-in' on their increased home equity and finally debt-financed purchases are generally encouraged by the prospect of rising future prices. Thus the data strongly supports the HH. Third, according to the model, the distribution of income does not play a leading role in explaining household debt accumulation. The negative long run coefficients on the top 1% share and the Gini coefficient are inconsistent with the ECH as well as with the ISH.

5.5 Consumption Expenditures and Income Inequality

The fact that the previous section did not find a positive relationship between income polarization and household liabilities does not necessarily falsify the ECH straight away. It might be the case that expenditure cascades are not debt-financed but only contributed to the decline of household saving rates. This section tests this interpretation by estimating private final consumption expenditures (C_{it}) as a function of disposable income (Y_{it}^D), income inequality (Q), property (PP_{it}) as well as stock prices (SP_{it}), the real long term interest rate (R_t), the ratio of people older than 65 in the population (OLD_{it}) and either the financial reform index from Abiad et al. (2008) (FIN_{it}) or the credit market regulation index published by the Fraser Institute ($CRED_{it}$) (subcategory 5A of the Economic Freedom of the World index). Lower case letters indicate log-transformations. Due to unit roots in the series, the data is first differenced prior to estimation:

$$\Delta c_{it} = \alpha_i + \beta_1 \Delta y_{it}^D + \beta_2 \Delta Q + \beta_3 \Delta pp_{it} + \beta_4 \Delta sp_{it} +$$

$$\beta_5 \Delta R_t + +\beta_6 \Delta OLD_{it} + \beta_7 \Delta s_{it} + \mu_{it} \qquad (19)$$

α_i is interpreted as country specific consumption growth rate unexplained by the other regressors. Columns (1) to (4) in Table 13 report the results from a fixed effects estimation of equation (19): The income elasticity is around 0.4 across all four specifications and statistically significant at the 1% level. Property price growth also exhibits a positive coefficient across all four specifications, again significant at the 1% level. The elasticity of 0.12 is in line with results reported in the literature (Ludwig & Sløk 2004; Paiella 2009). The stock price index yields a highly statistical significant elasticity of 0.02. Neither the top income share nor the Gini coefficient exhibit statistically significant coefficients. The estimated consumption function is therefore not compatible with the ECH. The coefficients on the real interest rate are not statistically significant across the specifications and neither are the two coefficients on the credit supply indices. The estimated constants correspond to autonomous consumption growth of 1% per period. This means the consumption model exhibits residual trends in consumption spending, unexplained by the other regressors, as was also the case in the debt equation.

Applying the mean group (MG) estimator to equation (19) allows to investigate the heterogeneity across countries in more detail. Results are reported in Table 13 columns (5) and (6) and are very close to the results of the fixed effects estimator which is expected as FE models yield consistent estimates of the average coefficient in static models. Therefore the income elasticities obtained by the MG estimator are around 0.4, the coefficients on residential property prices are somewhat higher at 0.14 and 0.18 but highly statistically significant. Also the stock price elasticities are very similar at 0.02. The most important difference is that the top income share coefficient is 0.84 and statistically significant at the 10% level. A look at the individual country results (reported in Table 7 in Appendix IV) reveals that there is only one country with a statistically significant top income share coefficient and that is Italy. However out of the 13 countries 9 exhibit positive top income share elasticities, although statistically not significant. A positive coefficient would be in line with the ECH, however due to the negative top income elasticity it would mean that rising inequality has a positive impact on spending but not on borrowing. So people would be willing to lower their saving rate in order to keep up with the Joneses but would not be willing to borrow in order to achieve that.

Column (7) contains the results from re-estimating equation (19) on a subsample which excludes Germany and Japan as the only two countries which did not experience a residential real estate boom in the period prior to 2007. The findings are very similar to those from the full sample specification. The disposable income elasticity is 0.29, the property price elasticity is 0.12 and the stock price elasticity is 0.02, all of them statistically significant at the 1% level. The top income share elasticity is not statistically significant. This means the consumption data does not support the ECH which would imply a positive impact of top income inequality on aggregate consumption growth due to status seeking spending.

Table 13: Consumption growth regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	FE, all countries				MG, all countries		FE, ex DE & JP
Δy^D	0.39***	0.39***	0.37***	0.35***	0.41***	0.35***	0.29***
	(0.09)	(0.09)	(0.09)	(0.09)	(0.06)	(0.06)	(0.06)
Δpp	0.12***	0.12***	0.12***	0.12***	0.14***	0.18**	0.12***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.09)	(0.02)
Δsp	0.02***	0.02***	0.02***	0.02***	0.01*	0.02***	0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)
$\Delta TOP1$	0.10		-0.19		0.84*		-0.12
	(0.10)		(0.14)		(0.43)		(0.12)
$\Delta GINI$		0.05		0.04		0.07	
		(0.08)		(0.08)		(0.23)	
ΔR	0.04	0.03	0.02	0.03	0.02	0.07	0.03
	(0.02)	(0.02)	(0.03)	(0.03)	(0.04)	(0.09)	(0.03)
ΔOLD	-0.09	-0.05	-0.21	-0.24	-0.89	-0.95	0.14
	(0.42)	(0.42)	(0.40)	(0.42)	(0.92)	(1.25)	(0.48)
Δfin	0.00	0.00					
	(0.01)	(0.01)					
$\Delta cred$			0.00	0.00	-0.02	-0.02	0
			(0.01)	(0.01)	(0.02)	(0.03)	(0.01)
$const$	0.01***	0.01***	0.01***	0.01***	0.01***	0.02***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
N	293	293	319	319	320	320	269
adj. R^2	0.45	0.45	0.45	0.44			0.53

Fixed effects (FE) and mean group (MG) estimators, dependent variable: Δc_{it}, robust standard errors clustered at the country level in brackets. Stars indicate statistical significance: * p<0.1, ** p<0.05, *** p<0.01. Lower case letters indicate variables are transformed by taking natural logarithms. Specifications (1) and (2) are based on data from 1980-2005, all other specifications based on data from 1980-2007. Specifications (1) to (6) include all 13 countries of the sample (AU, BE, CA, DE, FI, FR, IT, JP, NL, NO, SE, UK, US). Specification (7) includes countries which experienced increases of real residential property prices by more than 80% over the sample period (AU, BE, CA, FI, FR, IT, NL, NO, UK, US, SE).

The consumption function was also estimated in a partially restricted error correction form: allowing for heterogeneous short run and equilibrium adjustment effects but restricting the long run equilibrium condition to be homogeneous across countries:

$$\Delta c_{it} = \alpha_i + \beta_i^1 \Delta y_{it}^D + \beta_i^2 \Delta Q_{it} + \beta_i^3 \Delta pp_{it} + \beta_i^4 \Delta sp_{it} + \beta_i^5 \Delta R_{it} + \beta_i^6 \Delta OLD_{it} - \gamma_i \left(c_{it-1} - \beta^7 y_{it}^D - \beta^8 Q_{it} - \beta^9 pp_{it} - \beta^{10} sp_{it} - \beta^{11} R_{it} - \beta^{12} OLD_{it} \right) + \mu_{it} \quad (20)$$

Pooled Mean Group (PMG) results for equation (20) are presented in Table 14. In order to determine the lag structure of equation (20) a testing down procedure is used. First a fully specified model including one lag of all independent variables was estimated and then statistically insignificant short run effects were removed. As part of that process the credit market regulation index ($CRED_{it}$) was eliminated from the specification. The first two columns of Table 14 display the results based on the full set of countries and using only observations from the pre-crisis period (1980-2007). Specification (1) uses the top 1% income share and specification (2) the income Gini coefficient as indicators of income inequality. In both cases there is a statistically significant adjustment towards the estimated long run trend of about 20% per period. The estimated long run income elasticities are 0.68 and statistically different from 1 at the 1% level. An income coefficient less than 1 means that, while other factors are held constant, saving rates increase as income grows. This can be interpreted as additional evidence in support of the importance of house prices for household spending and borrowing decisions: The reason why saving rates and consumption to income ratios did not actually decline in most countries is because spending increased in the face of rising asset prices. The top 1% income share exhibits a negative coefficient but is statistically significant only at the 10% level. This finding contradicts the ECH because rising top income shares should trigger consumption spending further down in the income distribution and thus boost aggregate consumption. Real interest rates do not have a statistically significant impact on long run consumption spending, while the old age ratio has a statistically significant negative impact. This latter finding contrasts the standard life cycle hypothesis because the life cycle framework predicts that people dissave as they get older in order to consume. However given that people want to leave bequests or did not save enough for retirement this conclusion needs not to hold.

In columns (3) and (4) the specifications also include the post-crisis period. The adjustment speed towards the estimated long run relationship is lower (although the difference is not statistically significant) compared to the short sample in columns (1) and (2). The long run real estate and equity price elasticities are highly statistically significant and similar in magnitude. The top 1% income share exhibits a statistically significant negative long run elasticity and thus again falsifies the ECH. Interest rate elasticities (which were eliminated as part of the testing down procedure in specification 4) as well as the long run elasticities of the old age ratio are similar to the results obtained from specification (1) and (2). In columns (5) and (6) Germany and Japan are excluded from the sample due to the fact that they did not experience strong real estate price increases prior to the 2007 turmoil. Results are qualitatively very close to the results from specification (1) and (2), one important difference is the higher adjustment speed of 0.24 per annum.

Based on the growth regressions as well as the Error Correction Models two conclusions emerge. First, the importance of asset prices for household spending is confirmed. This is not only reflected in positive and highly significant long run asset price elasticities of consumption but also by the fact that income elasticities below unity imply increasing saving rates and declining consumption to income ratios. The fact that neither of those implications can be observed in the data points to the significant impact of asset price increases on consumption expenditures, providing support for the

HH. Second, the data does not support the ECH as an explanation for declining saving rates and strong consumption growth.

Table 14: Consumption, error correction models, pooled mean group (PMG) estimator

	(1)	(2)	(3)	(4)	(5)	(6)
sample	1980-2007		1980-2011		1980-2007, excl. DE & JP	
y^D_{t-1}	0.683***	0.687***	0.729***	0.621***	0.685***	0.672***
	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)
pp_t	0.088***	0.077***	0.076***	0.082***	0.088***	0.089***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
sp_t	0.050***	0.056***	0.038***	0.048***	0.049***	0.052***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
$TOP1_t$	-0.312*		-0.593***		-0.324*	
	(0.19)		(0.17)		(0.18)	
$GINI_t$		-0.185		-0.047		-0.112
		(0.14)		(0.11)		(0.14)
R_t	0.093	0.067	-0.066		0.088	0.159**
	(0.07)	(0.09)	(0.07)		(0.07)	(0.07)
OLD_t	-2.026***	-1.559***	-1.898***	-0.143	-2.046***	-1.822***
	(0.39)	(0.36)	(0.41)	(0.39)	(0.39)	(0.39)
	short run		short run		short run	
$adjustment$	-0.208**	-0.185***	-0.186***	-0.156***	-0.239**	-0.240***
	(0.08)	(0.06)	(0.07)	(0.05)	(0.10)	(0.09)
Δy^D_t	0.274***	0.270**	0.334***	0.294***	0.184**	0.174**
	(0.10)	(0.11)	(0.10)	(0.10)	(0.08)	(0.09)
Δpp_t	0.130***	0.109***	0.117***	0.118***	0.105***	0.097***
	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
Δsp_t		0.013***	0.012**	0.015***		0.012**
		(0.00)	(0.01)	(0.01)		(0.00)
$\Delta TOP1_t$			0.485**			
			(0.22)			
$constant$	0.501**	0.437***	0.393***	0.389***	0.570**	0.586***
	(0.20)	(0.15)	(0.15)	(0.14)	(0.23)	(0.23)
N	324	399	376	399	274	274
H0: $\beta_{y^D} = 1$	0.00	0.00	0.00	0.00	0.00	0.00

Error correction models with homogeneous long run effects and heterogeneous short run effects and adjustment speeds. Dependent variable: Δdh_t. Stars indicate statistical significance: * p<0.1, ** p<0.05, *** p<0.01. Lower case letters indicate variables are transformed by taking natural logarithms. Standard errors in brackets. The number of lags is determined via a testing down approach.

5.6 Summary of Chapter 5

The aim of this chapter is to fill a gap in the macroeconometric literature by taking the distribution of income as well as the evolution of residential real estate prices into account when analysing the determinants of household debt. These two factors are analysed separately in existing studies. A household debt as well as an auxiliary consumption function are estimated in order to assess the explanatory power of the three hypotheses under consideration. Results stemming from fixed effects regressions on logarithmic differenced data (growth regressions) as well as from autoregressive distributed lag models, confirm the results from the previous two chapters.

First, coefficients on residential real estate prices are highly statistically significant and positive in the estimated household debt and consumption equations. This result holds independent of which estimator is used or which sample or specification is analysed. Property prices emerge from assessing economic significance as the second biggest contributor to household debt in the period between 1995 and 2007 after disposable household income. Therefore, also macroeconomic data confirms the fundamental role of the housing market for household indebtedness and thus lend strong support to the housing hypothesis which is itself consistent with the ideas of credit constrained households, Minskian financing behaviour and spending behaviour anchored in stock flow norms as discussed in section 2.

The second result is that an increasingly polarized distribution of income, is negatively associated with household borrowing as well as consumption spending based on the estimated model. This casts doubt on the ECH and the ISH which both identify an increasingly unequal distribution of income in OECD countries as the main driver of household indebtedness. Also economic significance computations reveal that the direct impact of distributional shifts between 1995 and 2007 was rather small: According to the model the increase in the top 1% income share over that period reduced household borrowing by about 5%. So therefore the importance of either self- or other-regarding social norms, often found in models inspired by behavioural economics, is not confirmed by the data.

The third result is that the other factors for which the regression models controlled for exhibit coefficients which are in line with expectations from economic theory. Household disposable income for example contributes positively to household borrowing and emerges as the most important determinant in the 1995 and 2007 period. This indicates not only that as real incomes grow, also the standard of living increases but also that household borrowing positively contributes to economic growth by funding construction activities as well as consumption expenditures. In addition rising long term real interest rates negatively affect household borrowing and relaxing credit market regulations also enable credit to the household sector to expand. Interestingly however the results in this chapter suggest that an increasing share of people aged over 65 negativel contribute to household borrowing and consumption in the period 1980-2007. This finding is at odds with the life-cycle model which predicts consumption to increase and borrowing to decrease with age. Nevertheless effect size computations show that all of these factors only played a minor role in the household debt expansion between 1995 and 2013.

Two qualifications should be made explicit. First, the results presented here are based on panel analysis. Thus it cannot be rejected that for individual countries expenditure cascades may have played an important role. In particular for the USA there is microeconometric evidence that rising income inequality has had positive effects on consumption expenditures (Bertrand & Morse 2013) and household borrowing (Wildauer 2016; Carr & Jayadev 2015). We also do find evidence that inequality had a positive impact on consumption spending in Germany, which is also supported by microeconometric evidence (Drechsel-Grau & Schmid 2014). However, the results do suggest that this experience is not typical for OECD countries. Expenditure cascades do not seem to be a general mechanism. Second, these findings do not necessarily invalidate that households employ upward looking consumption. It may well be that this is how households reach their consumption decisions. However, banks may use collateral as the key variable in their lending decisions. Thus even if households' desired consumption is based on comparison with the richer Joneses, banks may look at the available collateral. Overall, property prices and the value of collateral seem to be the binding constraint which is of major importance in determining whether a credit expansion does happen. This does imply a word of caution towards the enthusiasm with which part of heterodox macroeconomics (Kapeller & Schütz 2014; Belabed et al. 2013) has embraced upward looking consumption norms.

6 Summary and Conclusion

The aim of this thesis is to compare three explanations of household sector borrowing and assess their consistency with the available data. Two hypotheses consider the rise in income inequality over the last 30 years as the primary reason for increased household indebtedness and therefore also as an explanation of the financial crisis of 2007/2008. First, the expenditure cascades hypothesis (ECH) argues that rising income inequality drove people into debt via status driven expenditures (‘keeping up with the Joneses’). Several authors argue on theoretical (Cardaci 2014; Frank et al. 2014; Kapeller & Schütz 2014; Ryoo & Kim 2014; Nikiforos 2015; Cynamon & Fazzari 2016) and empirical (Belabed et al. 2013; Bertrand & Morse 2013; Carr & Jayadev 2015; Perugini et al. 2016) grounds that expenditure cascades explain the rise in household debt and therefore identify it as a root cause of the financial crisis. Second, the income stagnation hypothesis (ISH) emphasizes that for a large proportion of the income distribution real incomes did not grow over the last decades. As a reaction to stagnant incomes households borrowed in order to maintain their living standards. Third, the housing hypothesis (HH) emphasizes the role of property prices for household borrowing. Rising real estate prices lead to demand for bigger mortgages relative to household income by (first time) buyers. Banks are willing to provide these mortgages because they rely on the property as seemingly save collateral. In addition rising prices create large capital gains for homeowners which they might ‘consume’ by taking on additional mortgages and the prospect of capital gains fuels debt-financed purchases in general. There is substantial empirical evidence that household borrowing (Bezemer & Zhang 2014; Borio 2014; Mian & Sufi 2011; Dynan & Kohn 2007; Haurin & Rosenthal 2006; Hurst & Stafford 2004) and spending (Cooper & Dynan 2016; Paiella 2009) are strongly influenced by real estate prices and that the resulting dynamics are key for financial stability. In order to compare these three hypotheses a special focus is put on the household sector in the United States because the US is the biggest economy in the world and the quality and the amount of available data, in particular the Survey of Consumer Finances (SCF), is unmatched. In addition to US survey data, a panel of 13 OECD countries[54] is analysed as well.

After discussing in chapter 2 how different schools of thought analyse household debt, the assessment of the two main hypotheses starts with a descriptive analysis of the SCF. The first result which emerges from the data is that the substantial increase in US household liabilities needs to be understood as the result of two effects. On the one hand, the proportion of US households holding any form of debt increased from 72.3% in 1989 to 77% in 2007 (‘more indebted households’). On the other hand, debt to income ratios increased for those households holding debt from 51% to 111% over the same period (‘households became more indebted’). Together both phenomena led to the rise of US household liabilities from \$4.4 trillion in 1989 to \$12.7 trillion in 2007, in 2013

[54] The countries included are: Australia, Belgium, Canada, Germany, Finland, France, Italy, Japan, Netherlands, Norway, Sweden, United Kingdom and the US.

prices. The second result is that, distinguishing household debt by type (real estate secured debt, instalment loans, credit card debt and other debt) as well as by the way it was used (real estate purchases and improvements, consumption, education), reveals the dominant role the housing market played for household borrowing. Real estate secured debt accounted for about 79% of total liabilities in 1998 and for 85% in 2007. In comparison, 78% of total liabilities in 1998 and 82% in 2007 were used for home purchases and improvements. This means that only a small but increasing proportion of real estate secured debt was used for other purposes than home purchases or improvements, \$173 billion in 2001 and \$276 billion in 2004. The third result of the descriptive analysis shows that there is no clear positive direct relationship between within group income inequality and household debt. Reference groups within which status comparison takes place are defined based on race and education in order to identify households which are likely to live in a similar area and engage within a similar milieu. The data only supports a direct relationship between within group inequality and indebtedness for black households but not for white households and especially not in the pre-crisis period between 2001 and 2007.

Chapter 4 moves from a purely unconditional descriptive analysis of SCF data to a regression framework. Because it is household borrowing and thus the change in debt rather than the stock of debt which is related to the flow of income in the current period or the decision to buy a property, an adequate empirical analysis requires a measure of household borrowing. In order to achieve this, the chapter develops a method which allows to infer borrowing from the stock of outstanding liabilities and credit history information such as the amount initially borrowed and the dates of taking on liabilities. The resulting measure of household borrowing is then used in a regression where the sample is split between borrowing households (those with positive changes in debt in the current period: $\Delta D_t > 0$) and non-borrowing households (households which either do not take on debt or are repaying already existing liabilities: $\Delta D_t \leq 0$). This split leads to an increased model fit and prevents estimating meaningless averages when the coefficients are substantially different across groups. This is the case for example with the estimated coefficients on income, real estate wealth and the stock of past liabilities. This finding represents the first important result which is that borrowing decisions are fundamentally different and need to be modelled separately from repayment decisions. The current literature on household borrowing does not pay attention to this issue. The second finding is that measures of real estate wealth and property prices exhibit statistically highly significant and positive coefficients across various specifications. This finding is very well in line with the results from the previous chapter and shows the importance of real estate transactions for household borrowing. The third finding is that an increase in within group income inequality, when reference groups are defined based on race and education as in chapter 3, is associated with increased household borrowing but only for homeowners. Therefore, the regression analysis points towards the importance of the distribution of income but the channel is conditional on homeownership. So while for the US case changes in the distribution of income had a significant impact on household borrowing decisions, the housing market and the availability of collateral in particular is identified by the data as the main driver and the binding constraint.

In chapter 5 a debt equation and an auxiliary consumption function are estimated for a panel of 13 OECD countries spanning from 1980 to 2007. Growth regressions provide the starting point: The time series enter the model after taking first differences. Due to the logarithmic transformation of the monetary variables the resulting logarithmic differences serve as proxies for the growth rates of these series. In a second step the dynamic restrictions which are implicitly imposed when using differenced data only, are relaxed and autoregressive distributed lag models (ARDL) are estimated. While allowing a richer dynamic structure is desirable from a time series perspective, the fact that the model is estimated for a panel means that with more lags also more parameters are restricted to be the same across countries. In order to partially relax this assumption the Pooled Mean Group (PMG) estimator is applied, which estimates an error correction model (and thus a reparametrized ARDL) which allows short run coefficients to be country specific and only restricts long run elasticities to be equal across countries. This approach represents a pragmatic compromise between allowing for cross country heterogeneity while limiting the number of parameters which need to be estimated. The most robust result from these panel regressions is that real estate prices have a positive and highly statistically significant impact on household debt and also on household consumption expenditures. This result holds across specifications, sample periods and estimation methods. This means that chapter 5, in line with the previous two chapters, supports the notion that the housing market plays a fundamentally important role for household borrowing. The second finding is that an increased polarization of the income distribution negatively impacts household borrowing and household consumption spending in this sample of OECD countries. In addition economic significance computations based on the panel average, reveal that changes in the top 1% income share contributed to a decrease in household borrowing by only 5% between 1995 and 2007, in contrast to the positive contribution of real estate prices of 43% over the same period.

Based on these findings I draw four major conclusions: The first is that the data is overwhelmingly consistent with the housing hypothesis. The perspective that household debt is mainly driven by real estate transactions and property prices seems to be a very good way of looking at and thinking about household debt dynamics. The reason why property plays the dominant role in determining household liabilities is that acquiring it represents the biggest type of expenditure for most households and it can be used as collateral. If household debt is mainly driven by inflated residential real estate prices, then introducing caps on loan-to-value ratios as well as increasing the supply of (public) housing could be a policy mix which has the potential to avoid excessive household debt accumulation while the cost of forgone output growth is minimal because home building is not reduced. This asset-focused view of household debt is consistent with the idea of credit constrained households, a Minskian idea of optimism facilitated through asset prices driving borrowing and expenditure decisions anchored in stock flow norms. However it is in sharp contrast to the textbook life-cycle perspective and thus leads to the second conclusion: Borrowing in order for consumption smoothing is dwarfed by borrowing for real estate purchases and improvements. In addition, households but also central bankers (Greenspan 2004) are not able to distinguish between permanent and transitory changes in asset values. In contrast temporary increases in real estate prices, which might last for years or decades, can trigger unsustainable household debt accumulation. Thus a distinction between transitory and temporary asset values seems not very useful in analysing household behaviour. Third, the distribution of income cannot be seen as the main and dominant driver of household indebtedness. The reason is that housing variables emerge as the more important explanatory variables in every analysis carried out in this thesis and even in cases when there is some evidence that rising income polarisation leads to higher household indebtedness, as is the case in Chapter 4, these findings are conditional on home ownership. This does imply a word of caution towards the enthusiasm with which parts of heterodox macroeconomics (Cardaci 2014; Frank et al. 2014; Kapeller & Schütz 2014; Ryoo & Kim 2014; Nikiforos 2015; Cynamon & Fazzari 2016; Belabed et al. 2013) has embraced upward looking consumption norms. This does not mean that expenditure cascades are not relevant at all. If they describe the behaviour of a smaller subgroup of households which over accumulates debt, it might be that the defaults in this subgroup have the potential to trigger knock on effects and develop into a full scale financial crisis. It might also be the case that upward looking consumption norms are a valid factor in forming household demand but household borrowing is constrained by (real estate) collateral. In both cases however it would be misleading to interpret soaring household debt as a direct result of rising income inequality. Fourth, the fact that different analyses yield different results regarding the impact of the distribution of income on household borrowing, points out that further research is needed. It seems that the interaction of the income distribution with household debt and the macroeconomy is not yet fully understood and it might be the case that these effects are fundamentally different for different countries. The necessity of further research about the role of the distribution of income derives not only from the results obtained as part of this thesis but also from the fact that even in the case of the US, positive (Carr & Jayadev 2015; Bertrand & Morse 2013) as well as negative (Coibion et al. 2016) effects of the income distribution on household borrowing and spending are found.

Overall the results of this thesis contribute to the analysis and understanding of household borrowing decisions in three important ways. First, the thesis provides a comprehensive analysis of household liabilities which takes the distribution of income as well as asset prices into account. The current literature usually treats these two factors separately and does not analyse them in a joint framework. Second, the first two chapters of this thesis are the first attempt to use the high-quality data provided by the Survey of Consumer Finances to investigate the impact of shifts in the income distribution on household borrowing. The PSID and the CEX which are used by existing studies of distributional impact on borrowing, do not have the SCF's level of technical sophistication when it comes to dealing with non-observation and differential non-response problems. Therefore they almost certainly underestimate the degree of income inequality. For studies which try to analyse the impact of the income distribution this represents a major shortcoming. Third, chapter 4 provides some methodological insights for investigating household borrowing based on survey data. With cross sectional data, borrowing and non-borrowing households need to be analysed separately. The benefits of a separate analysis materialize in an increased model fit as well as easier interpretable coefficients as is demonstrated by differently signed income effects in the borrowing and non-borrowing sample. In addition since property purchases are the single most important reason to borrow for households, controlling for such purchases is important. This means that the analysis of household borrowing requires either panel data which automatically provides information on purchases (as long as assets are observed) or in a cross section context information about the timing of purchases is needed in order to be able to explain the large spikes in household borrowing related to property acquisitions. Finally, because property purchases lead to sudden and large increases in household debt, growth rates in general and logarithmic differences in particular are not very informative measures of household borrowing. Identical growth rates can result from households engaging in very different behaviour. Using growth rates also eliminates borrowers with no previous debt in the sample.

The policy implications of these findings are substantial. First, real estate price dynamics are the primary factor driving household sector liabilities and thus need to be closely monitored by those policy institutions which are responsible for maintaining financial stability. This raises non-trivial questions for the conduct of monetary policy in situations when hitting a growth target for the economy and maintaining financial stability are in conflict due to mounting household liabilities. It also raises questions about the responsibilities of fiscal policies and to what extent fiscal measures should be used to ensure financial stability. For example large investment programs in the stock of residential real estate could play an important part in an effort to keep real estate price rallies in check. Similarly investing in and upgrading the existing transport infrastructure can ease the pressure on housing markets with local limits for supply expansions, such as the centres of major cities. Second, the lack of data is a fundamental problem for studying the links between the distribution of income, asset prices and household liabilities. In the first wave of the ECB's Household Finance and Consumption Survey only nine out of fifteen participating countries made an effort to oversample wealthy households and only Spain and France relied on individual tax information to do so (ECB 2013). One specific way to enhance future research would be to conduct high quality surveys relying on oversampling methods using household level tax data. More effort and dedication from local central banks and tax authorities is needed. In addition central banks should provide more resources for collecting data in order to enable bigger samples and, most importantly, shorter intervals of data collection. Three year intervals, as is the case with the SCF and the HFCS, leave substantial gaps which would be unimaginable in the case of GDP or inflation data.

Based on the results presented in this thesis several routes for future research emerge. First, if residential real estate prices are the main driver of household indebtedness, this poses the question of how much did this debt accumulation contribute to economic growth by financing residential investment expenditures. There is some evidence that household sector debt accumulation was a major driver of economic growth prior to the 2008 crisis (Stockhammer & Wildauer 2016) and thus policies which are aimed at preventing household debt bubbles should be designed in such a way that they do not interfere with the supply of housing. The goal would be to expand housing supply while keeping prices in check. Government policies such as public housing as well as moving away from the public's glorification of the single family house as the highest standard of living. To what extent such changes have the potential to stabilize housing markets and which measures are best suited to deliver them provides future research opportunities. Second, even though this work shows the relevance of stocks such as the value of residential real estate for household decision making (borrowing), the results do not provide general rules how these stocks influence decision making. Do households follow stock-flow norms as pointed out by Godley and Lavoie (2007, p.75)? If they do, are these norms stable over time? Understanding such relationships could be highly valuable for forecasting and anticipating how an economy develops over the near future.

Appendix I

Table I-1: Average Top1% incomes by group

	1989	1992	1995	1998	2001	2004	2007	2010	2013
no-col | w	$633	$427	$644	$762	$735	$672	$887	$546	$675
no-col | b	$185	$145	$167	$220	$391	$163	$358	$183	$195
col | w	$3,075	$1,330	$1,868	$2,243	$3,588	$2,737	$3,865	$2,511	$3,237
col | b	$519	$419	$397	$582	$557	$1,974	$996	$414	$473

Average incomes in thousands of 2013 US Dollars. Author's own calculation based on SCF waves 1989-2013.

Figure I-1: Aggregate disposable income and debt

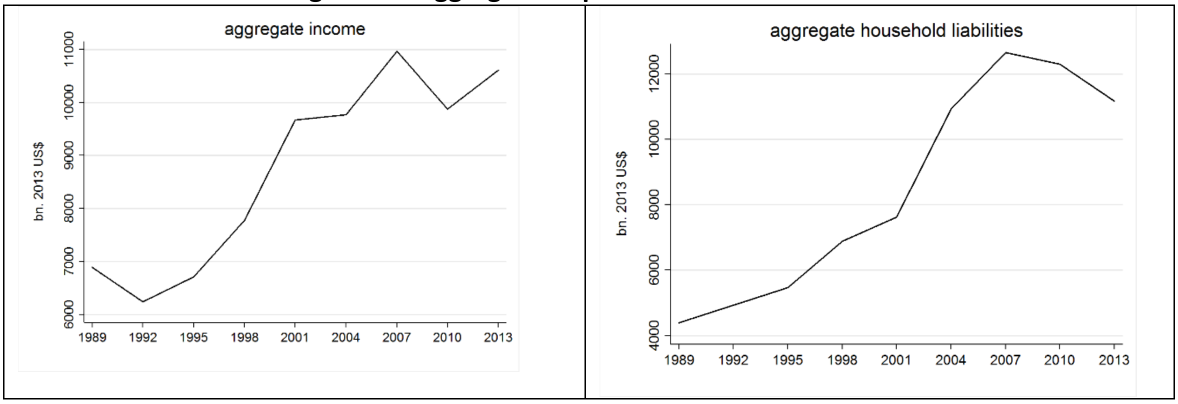


Figure I-2: Consumption borrowing relative to disposable income

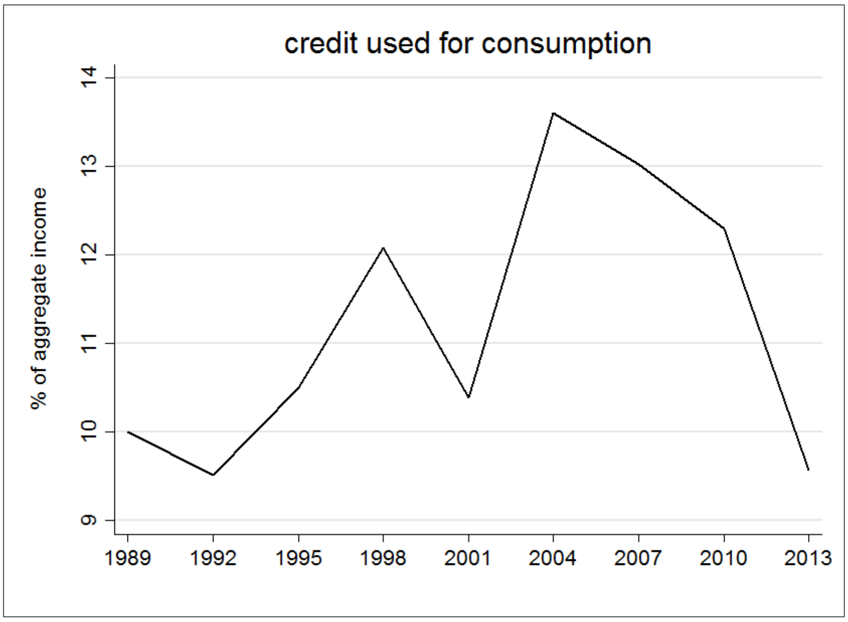


Figure I-3: A Debt by race and education, top 5% share, median ratios

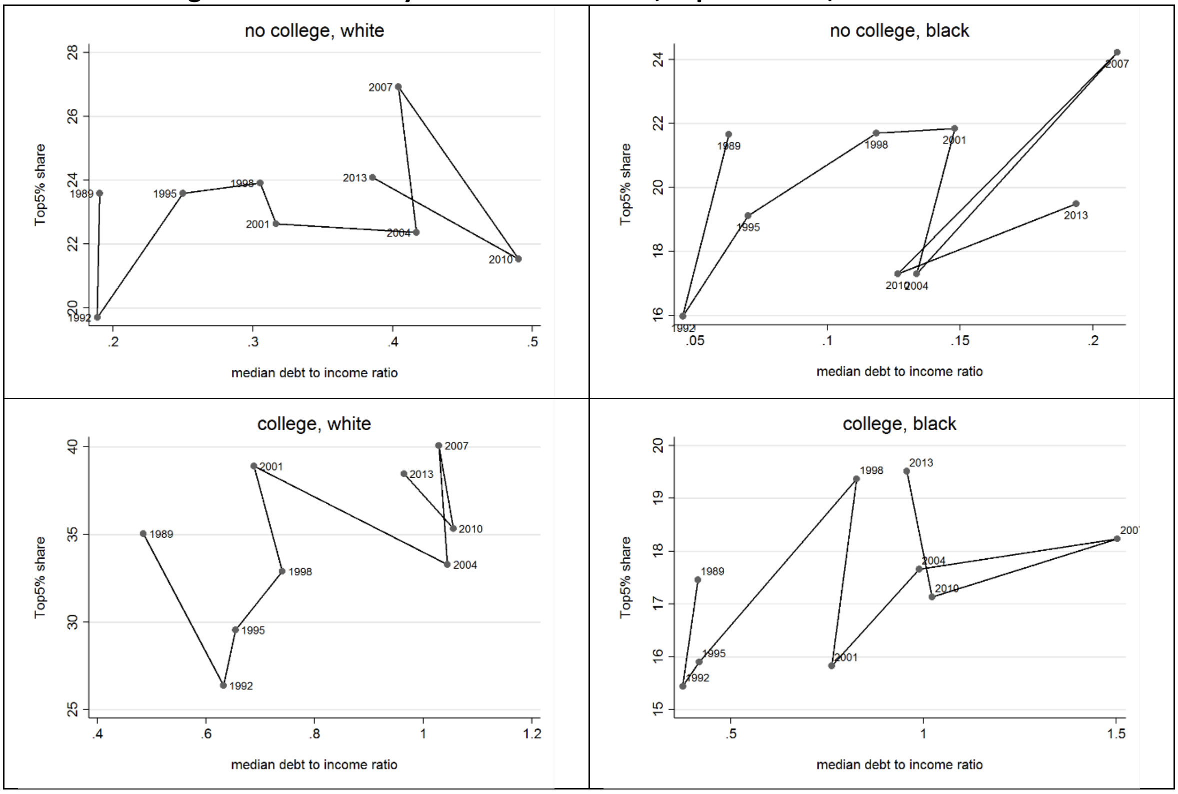


Figure I-4: Debt by race and education, top 1% share, absolute values

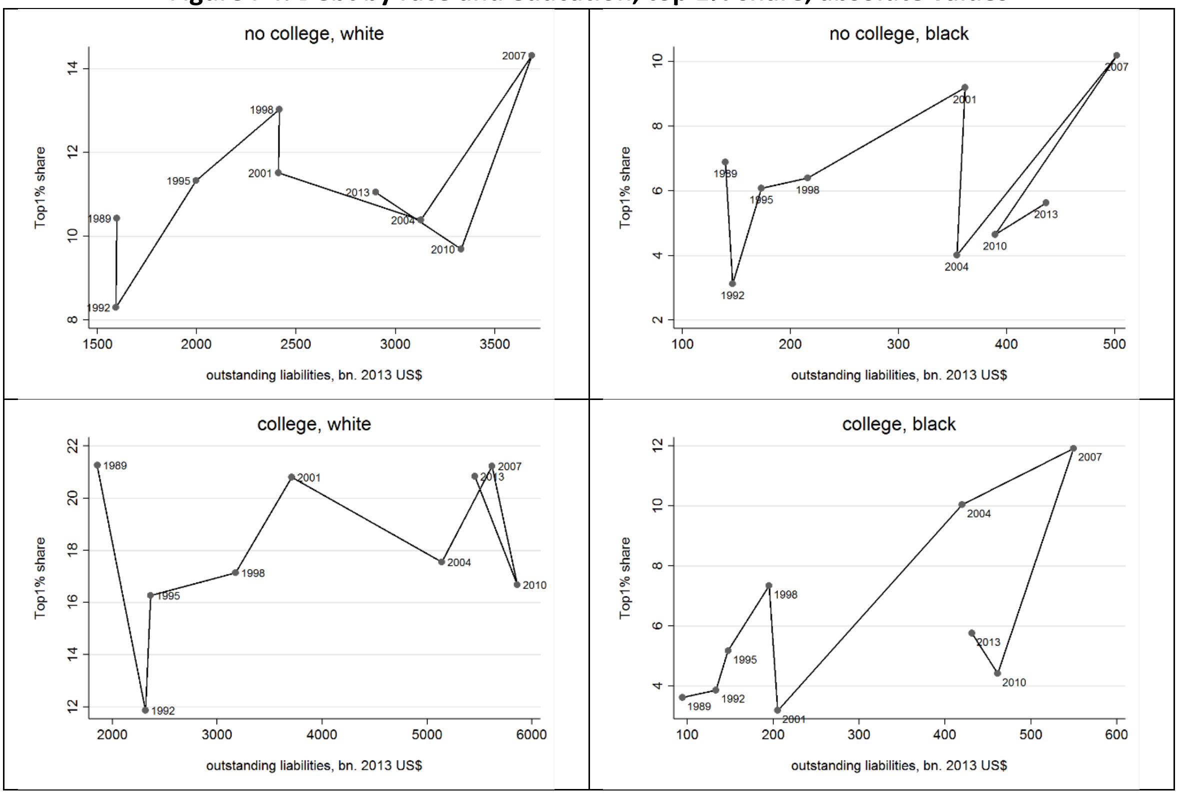
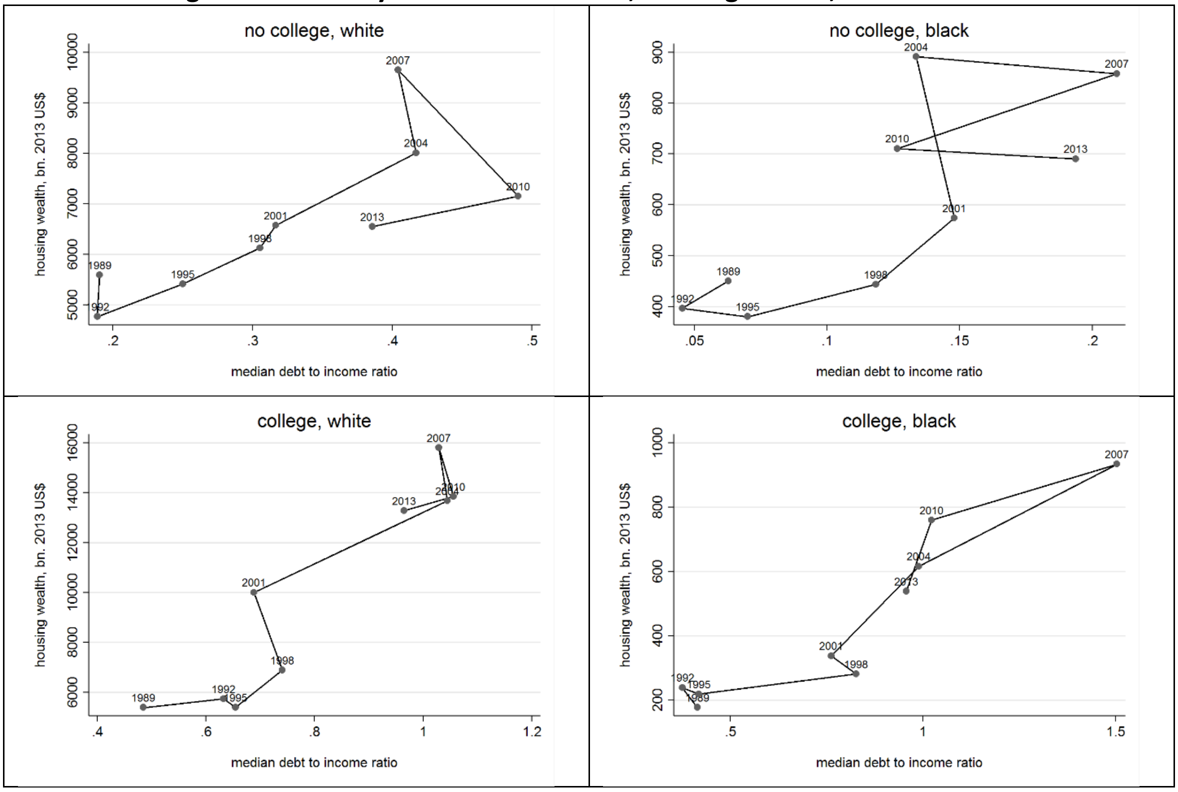
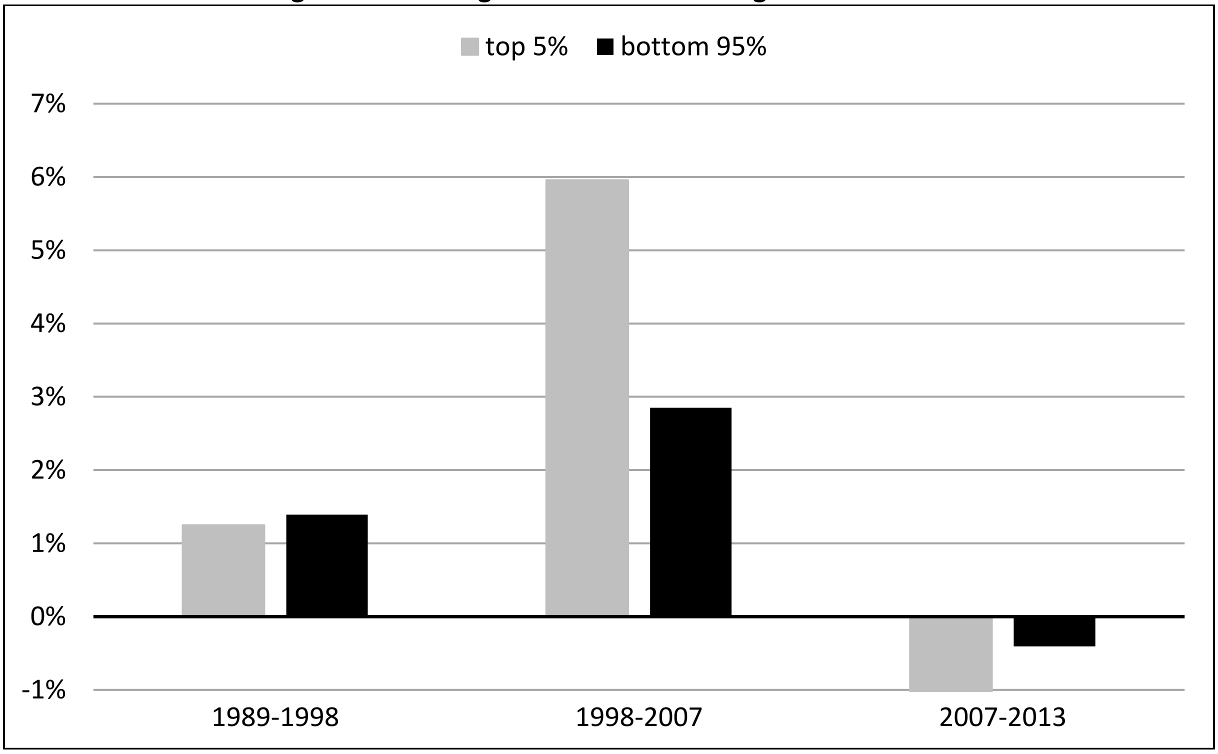


Figure I-5: Debt by race and education, housing wealth, median ratios



Appendix II

Figure II-1: Average annual real income growth rates



Source: own computations based on SCF waves 1989 to 2013.

Figure II-2: Decision tree, first mortgage on primary residence

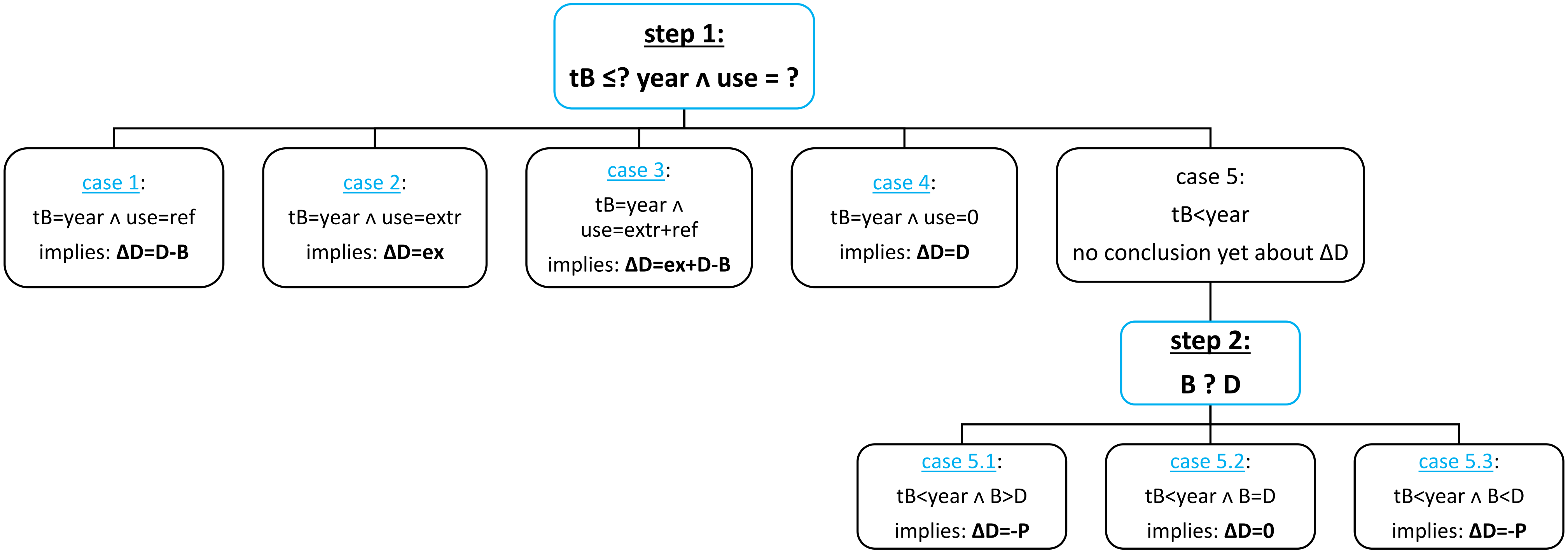


Table II-1: Variable description

Variable	Description	SCF code/name
ΔD_{it}	Household borrowing defined as the change in outstanding liabilities within year t. The way the measure is constructed is discussed in section 3.2 and Appendix B.	
D_{it-1}	Outstanding liabilities at the beginning of the period, derived from the definition: $D_{it-1} \equiv D_{it} - \Delta D_{it}$.	
Y_{it}	Total gross income as reported by the household. Including realised capital gains. Before taxes and other deductions.	Corresponds to the variable 'income' in the summary dataset.
HW_{it}	Value of residential real estate, minus the value of residential real estate purchases in the current period and minus any capital gains on these purchases. Thus HW_{it} corresponds to the value of residential real estate the household owned at the beginning of the period.	Residential real estate corresponds to the variables 'houses' and 'oresre' in the summary dataset. For the detailed computations see Stata code.
REP_{it}	Value of real estate purchases in the current period, excluding capital gains on these purchases.	
FW_{it}	Total value of financial assets. Includes: checkings and savings accounts, money market accounts, certificates of deposits, directly held mutual funds, stocks, bonds, quasi liquid pension accounts, savings bonds, cash in life-insurance products, other managed assets and other financial assets.	Coded 'fin' in the summary dataset.
$\tilde{Y}$	Definitions of relative income measures are discussed in section 4.1.	
age_{it}	Age of the household head.	Coded 'age' in the summary dataset.
$kids_{it}$	Dummy variable, 1 indicating the presence of children.	Based on variable 'kids' from summary dataset.
$college_{it}$	Dummy variable, 1 indicating the household head obtaining a college degree.	Based on variable 'edcl' from summary dataset.
$black_{it}$	Dummy variable, 1 indicating household head self-identified as black.	Based on variable 'race' from summary dataset.
$working_{it}$	Dummy variable, 1 indicating household head is part of the labour force.	Coded 'lf' in summary dataset.
$lowinc_{it}$	Dummy variable, 1 indicating household income is lower than in normal year.	Based on X7650 in full dataset.
$not\ married_{it}$	Dummy variable, 1 indicating household head is not married or living with a partner.	Coded 'married' in summary dataset.
$turndown_{it}$	Dummy variable, 1 indicating that in the past five years the household had been turned down when applying for credit .	Coded 'turndown' in the summary dataset. Based on X407 in full dataset.

Table II-2, Part1: Complete version of Table 4

	(1)	(2)	(3)	(4)	(5)
sample	ΔD/Y>0	ΔD/Y≤0	full	ΔD/Y>0	ΔD/Y≤0
Y~ cut-off	99th perc	99th perc	99th perc	99th perc	99th perc
Y~ reference definition	av. inc.	av. inc.	av. inc.	av. inc.	av. inc.
Y	-0.314	0.042	0.001	-0.314	0.042
	(0.024)***	(0.003)***	(0.005)	(0.023)***	(0.003)***
Y~	0.216	0.009	0.056		
	(0.080)***	(0.006)	(0.019)***		
HW0 # Y~				0.089	0.004
				(0.115)	(0.006)
HW1 # Y~				0.289	0.010
				(0.085)***	(0.007)
HW	0.086	-0.002	0.006	0.090	-0.003
	(0.017)***	(0.001)	(0.003)*	(0.017)***	(0.001)**
dHW	-1.004	0.006	-0.077	-3.889	-0.074
	(0.200)***	(0.016)	(0.035)**	(1.526)**	(0.108)
REP	0.659	0.011	0.539	0.658	0.012
	(0.047)***	(0.011)	(0.038)***	(0.048)***	(0.011)
dREP	-6.506	-0.150	-5.077	-6.507	-0.153
	(0.563)***	(0.144)	(0.435)***	(0.576)***	(0.144)
FW	0.008	-0.001	0.000	0.008	-0.001
	(0.006)	(0.001)	(0.001)	(0.006)	(0.001)
dFW	0.228	-0.006	0.038	0.188	-0.004
	(0.077)***	(0.007)	(0.011)***	(0.081)**	(0.006)
D_{t-1}	-0.005	-0.024	-0.024	-0.007	-0.024
	(0.009)	(0.001)***	(0.002)***	(0.009)	(0.001)***
dD_{t-1}		0.148	0.215		0.149
		(0.005)***	(0.022)***		(0.005)***
kids	0.035	-0.001	0.011	0.039	-0.001
	(0.023)	(0.002)	(0.004)**	(0.022)*	(0.002)
age	-0.007	-0.001	-0.005	-0.007	-0.001
	(0.005)	(0.000)	(0.001)***	(0.005)	(0.000)
age^2	0.000	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)***	(0.000)	(0.000)
N	2,229	14,270	16,510	2,229	14,270
av. R^2	0.65	0.27	0.51	0.65	0.27

Table II-2, Part 2: Complete version of Table 4

	(1)	(2)	(3)	(4)	(5)
sample	ΔD/Y>0	ΔD/Y≤0	full	ΔD/Y>0	ΔD/Y≤0
Y~ cut-off	99th perc	99th perc	99th perc	99th perc	99th perc
Y~ reference definition	av. inc.	av. inc.	av. inc.	av. inc.	av. inc.
college	-0.269	-0.012	-0.073	-0.080	-0.016
	(0.111)**	(0.008)	(0.025)***	(0.163)	(0.009)*
HW1#college				-0.297	0.006
				(0.147)**	(0.011)
black	0.218	-0.002	0.044	-0.038	0.000
	(0.100)**	(0.007)	(0.023)*	(0.153)	(0.007)
HW1#black				0.443	-0.012
				(0.149)***	(0.008)
college#black	0.438	0.008	0.094	0.509	-0.006
	(0.098)***	(0.007)	(0.022)***	(0.134)***	(0.010)
HW1#collge#black				-0.145	0.029
				(0.211)	(0.015)*
working	-0.008	-0.001	-0.001	-0.003	-0.001
	(0.048)	(0.003)	(0.008)	(0.047)	(0.003)
lowinc	0.039	-0.017	-0.005	0.039	-0.016
	(0.030)	(0.003)***	(0.007)	(0.029)	(0.003)***
not married	-0.031	0.011	-0.007	-0.033	0.011
	(0.022)	(0.002)***	(0.005)	(0.023)	(0.002)***
turndown	0.035	-0.005	0.023	0.030	-0.004
	(0.023)	(0.003)*	(0.007)***	(0.023)	(0.003)
dum1995	-0.035	0.005	0.011	-0.035	0.005
	(0.029)	(0.003)*	(0.007)	(0.029)	(0.003)*
dum2001	0.013	0.001	0.003	0.013	0.001
	(0.032)	(0.003)	(0.007)	(0.033)	(0.003)
dum2004	0.063	0.004	0.023	0.062	0.003
	(0.032)*	(0.002)	(0.007)***	(0.033)*	(0.002)
dum2007	-0.074	0.003	-0.008	-0.075	0.003
	(0.042)*	(0.003)	(0.009)	(0.044)*	(0.003)
constant	0.758	-0.555	-0.669	2.607	-0.480
	(1.176)	(0.086)***	(0.269)**	(1.710)	(0.093)***
N	2,229	14,270	16,510	2,229	14,270
av. R^2	0.65	0.27	0.51	0.65	0.27

Table II-3, Part 1: Owner vs non-owner split

	(1)	(2)	(3)	(4)
sample	ΔD/Y>0	ΔD/Y>0	ΔD/Y≤0	ΔD/Y≤0
Y~ cut-off	99th perc	99th perc	99th perc	99th perc
Y~ definition	av. inc.	av. inc.	av. inc.	av. inc.
restriction	owners	non-owners	owners	non-owners
Y	-0.240	-0.427	0.051	0.019
	(0.025)***	(0.045)***	(0.003)***	(0.003)***
Y~	0.291	-0.008	-0.001	-0.001
	(0.106)***	(0.133)	(0.008)	(0.006)
HW	0.096		-0.006	
	(0.015)***		(0.001)***	
REP	0.475	0.840	0.020	-0.002
	(0.057)***	(0.062)***	(0.017)	(0.003)
dREP	-4.770	-8.405	-0.263	0.014
	(0.656)***	(0.783)***	(0.230)	(0.037)
FW	-0.004	0.010	0.000	-0.001
	(0.006)	(0.011)	(0.001)	(0.001)
dFW	-0.038	0.291	-0.002	0.006
	(0.125)	(0.111)**	(0.012)	(0.005)
D_{t-1}	-0.011	-0.001	-0.024	-0.022
	(0.009)	(0.014)	(0.001)***	(0.001)***
dD_{t-1}			0.158	0.149
			(0.007)***	(0.008)***
kids	0.020	0.009	-0.003	-0.002
	(0.022)	(0.042)	(0.003)	(0.002)
age	-0.006	-0.013	-0.002	0.001
	(0.007)	(0.009)	(0.001)***	(0.000)**
age^2	0.000	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)***	(0.000)**
N	1,503	718	10,212	3,667
F-stat	17	98	326	89

Table II-3, Part 2: Owner vs non-owner split

	(1)	(2)	(3)	(4)
sample	ΔD/Y>0	ΔD/Y>0	ΔD/Y≤0	ΔD/Y≤0
Y~ cut-off	99th perc	99th perc	99th perc	99th perc
Y~ definition	av. inc.	av. inc.	av. inc.	av. inc.
restriction	owners	non-owners	owners	non-owners
college	-0.356	0.001	-0.002	0.003
	(0.142)**	(0.186)	(0.011)	(0.008)
black	0.452	-0.116	-0.023	-0.008
	(0.125)***	(0.169)	(0.009)**	(0.007)
college#black	0.304	0.412	0.019	0.003
	(0.172)*	(0.129)***	(0.010)*	(0.006)
working	0.019	-0.084	0.003	-0.005
	(0.045)	(0.102)	(0.004)	(0.003)
lowinc	0.101	-0.010	-0.017	-0.005
	(0.038)***	(0.044)	(0.004)***	(0.003)
not married	-0.037	0.005	0.014	0.003
	(0.025)	(0.043)	(0.003)***	(0.002)
turndown	0.002	0.134	-0.007	-0.001
	(0.026)	(0.039)***	(0.003)**	(0.002)
dum1995	-0.028	-0.061	0.004	0.004
	(0.028)	(0.062)	(0.004)	(0.004)
dum2001	-0.018	0.049	0.002	0.004
	(0.035)	(0.068)	(0.003)	(0.004)
dum2004	0.005	0.113	0.005	0.004
	(0.035)	(0.060)*	(0.003)	(0.004)
dum2007	-0.078	-0.036	0.009	0.001
	(0.051)	(0.075)	(0.004)**	(0.004)
constant	-1.887	5.152	-0.456	-0.206
	(1.483)	(2.008)**	(0.115)***	(0.083)**
N	1,503	718	10,212	3,667
F-stat	17	98	326	89

Table II-4, Part 1: Robustness check grouping, borrowing sample

	(1)	(2)	(3)	(4)
sample	ΔD/Y>0	ΔD/Y>0	ΔD/Y>0	ΔD/Y>0
Y~ definition	avtop1	avtop1	head	head
grouping	edu-race	region	edu-race	region
Y	-0.291	-0.295	-0.526	-0.454
	(0.041)***	(0.040)***	(0.129)***	(0.105)***
HW0 # Y~	0.092	-0.056	-0.655	-0.407
	(1.150)	(0.145)	(0.369)*	(0.311)
HW1 # Y~	0.109	-0.021	-0.808	-0.657
	(1.188)	(0.113)	(0.345)**	(0.298)**
HW	0.088	0.084	0.076	0.077
	(0.021)***	(0.022)***	(0.021)***	(0.020)***
REP	0.540	0.546	0.496	0.501
	(0.056)***	(0.056)***	(0.054)***	(0.054)***
FW	0.018	0.019	0.015	0.016
	(0.008)**	(0.009)**	(0.008)*	(0.008)*
D_{t-1}	0.007	0.008	0.007	0.008
	(0.010)	(0.010)	(0.009)	(0.009)
region 2	0.117	0.125	0.108	0.088
	(0.058)*	(0.064)*	(0.058)*	(0.061)
region 3	0.131	0.123	0.122	0.090
	(0.059)**	(0.060)*	(0.056)**	(0.059)
region 4	0.055	0.045	0.044	-0.048
	(0.075)	(0.070)	(0.066)	(0.080)
region 5	0.065	0.057	0.053	0.023
	(0.068)	(0.076)	(0.065)	(0.068)
region 6	0.116	0.122	0.106	0.088
	(0.067)	(0.072)	(0.065)	(0.065)
region 7	0.079	0.067	0.059	-0.002
	(0.068)	(0.077)	(0.066)	(0.071)
region 8	0.055	0.022	0.041	0.007
	(0.064)	(0.094)	(0.062)	(0.068)
region 9	0.048	0.060	0.038	0.026
	(0.064)	(0.074)	(0.061)	(0.061)
dum1995	-0.037	-0.057	-0.058	-0.055
	(0.237)	(0.046)	(0.028)**	(0.029)*
constant	2.182	4.358	6.458	5.574
	(16.317)	(2.192)*	(1.659)***	(1.390)***
N	915	914	1,007	1,007

Dependent variable: ΔD/Y. All $ valued independent variables are subject to the inverse hyperbolic since transformation. Coefficients are estimated by OLS using probability weights. Bootstrapped standard errors are obtained by re-estimating the regression 999 times using a set of 999 replicate weights. Stars indicate 1% (***), 5% (**) and 10% (*) significance levels. Full set of results including missing household characteristics can be obtained upon request.

Table II-4, Part 2: Robustness check grouping, non-borrowing sample

	(5)	(6)	(7)	(8)
sample	ΔD/Y≤0	ΔD/Y≤0	ΔD/Y≤0	ΔD/Y≤0
Y~ definition	avtop1	avtop1	head	head
grouping	edu-race	region	edu-race	region
Y	0.040	0.040	0.015	0.013
	(0.003)***	(0.003)***	(0.005)***	(0.005)***
HW0 # Y~	0.008	0.035	-0.039	-0.057
	(0.016)	(0.052)	(0.017)**	(0.014)***
HW1 # Y~	0.011	-0.002	-0.120	-0.140
	(0.014)	(0.041)	(0.017)***	(0.017)***
HW	-0.006	-0.007	-0.011	-0.010
	(0.002)***	(0.002)***	(0.002)***	(0.002)***
REP	-0.008	-0.003	-0.002	-0.001
	(0.005)	(0.005)	(0.004)	(0.004)
FW	0.001	0.001	0.000	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
D_{t-1}	-0.025	-0.026	-0.026	-0.026
	(0.001)***	(0.001)***	(0.001)***	(0.001)***
region 2	-0.011	-0.011	-0.011	-0.015
	(0.008)	(0.008)	(0.007)	(0.008)*
region 3	0.005	0.003	0.001	-0.005
	(0.008)	(0.009)	(0.008)	(0.008)
region 4	-0.006	-0.008	-0.009	-0.025
	(0.009)	(0.010)	(0.009)	(0.009)**
region 5	-0.010	-0.016	-0.018	-0.023
	(0.011)	(0.010)	(0.010)*	(0.009)**
region 6	-0.001	-0.002	-0.004	-0.007
	(0.009)	(0.009)	(0.009)	(0.009)
region 7	-0.005	-0.009	-0.010	-0.019
	(0.010)	(0.010)	(0.010)	(0.010)*
region 8	0.009	0.002	0.001	-0.005
	(0.010)	(0.010)	(0.009)	(0.009)
region 9	0.006	0.008	0.007	0.006
	(0.012)	(0.011)	(0.010)	(0.010)
dum1995	0.005	0.003	0.000	-0.001
	(0.005)	(0.007)	(0.003)	(0.003)
constant	-0.525	-0.871	-0.118	-0.078
	(0.241)**	(0.696)	(0.066)*	(0.057)
N	5,460	4,919	6,415	6,415

Dependent variable: ΔD/Y. All $ valued independent variables are subject to the inverse hyperbolic since transformation. Coefficients are estimated by OLS using probability weights. Bootstrapped standard errors are obtained by re-estimating the regression 999 times using a set of 999 replicate weights. Stars indicate 1% (***), 5% (**) and 10% (*) significance levels. Full set of results including missing household characteristics can be obtained upon request.

Table II-5: Effect Size based on average income of top 1% of households within edu-race groups

	total	Y	HW0#Y~	Y~ + HW	REP	FW	D	time	kids
1998	0.77%	-3.26%	-0.51%	5.86%	0.62%	0.67%	-1.25%	0.10%	0.05%
2001	1.10%	-2.29%	-0.14%	6.16%	1.12%	0.40%	-3.16%	0.35%	0.05%
2004	6.99%	-8.62%	1.19%	11.21%	6.80%	1.67%	-3.65%	1.39%	0.13%
2007	2.32%	-2.86%	-0.33%	8.66%	3.46%	0.53%	-4.65%	-0.61%	0.05%

	age	college	black	college#black	working	lowinc	not married	turndown
1998	-0.82%	-0.64%	-0.09%	0.05%	-0.01%	0.00%	0.01%	0.02%
2001	-0.96%	-0.58%	0.18%	-0.04%	-0.03%	-0.03%	0.05%	0.01%
2004	-2.00%	-1.25%	0.12%	0.08%	-0.04%	-0.07%	0.02%	0.01%
2007	-1.30%	-0.72%	0.08%	0.02%	-0.03%	-0.05%	0.05%	0.00%

Table II-6: Effect Size based on average income of top 5% of households within edu-race groups

	total	Y	HW0#Y~	Y~ + HW	REP	FW	D	time	kids
1998	0.97%	-1.28%	-0.76%	3.72%	0.35%	0.27%	-1.04%	0.36%	0.02%
2001	0.75%	-0.72%	-0.64%	4.26%	0.33%	0.26%	-2.29%	0.48%	0.03%
2004	5.62%	-4.08%	1.39%	6.70%	4.62%	0.83%	-2.99%	0.99%	0.05%
2007	2.02%	0.10%	-0.44%	4.62%	2.43%	0.20%	-3.53%	-0.22%	0.02%

	age	college	black	college#black	working	lowinc	not married	turndown
1998	-0.57%	-0.14%	-0.05%	0.07%	-0.02%	0.00%	0.02%	0.01%
2001	-0.83%	-0.17%	0.12%	-0.07%	-0.03%	-0.03%	0.04%	0.02%
2004	-1.55%	-0.39%	0.05%	0.10%	-0.07%	-0.06%	0.03%	0.00%
2007	-0.99%	-0.20%	0.05%	0.00%	-0.03%	-0.05%	0.06%	0.00%

Appendix III

This appendix describes how the change in outstanding liabilities was defined for the other debt categories used in this paper beyond the first mortgage on the primary residence which was discussed in section 2. These other debt categories are second and third mortgages on the primary residence, mortgages on other residential properties, consumer loans, car and vehicle loans, education loans, other loans for property purchase and home improvements. It is noted that the SCF also collects information on credit card debt, credit lines (including home equity lines), loans against land contracts and loans against pension plans. These categories have been omitted as the information on when these liabilities were taken on is not available and therefor inferences about how much outstanding amounts changed within the year of the survey cannot be made.

6.1 Second and third mortgage on primary residence

Table III-1: Changes in the outstanding amount of the second and third mortgage on the primary residence

case	step 1	step 2	definition	N	interpretation
1	tB = year		ΔD = D	25	Assumption is that second and third mortgages are not refinanced but new debt.
2.1	tB < year	B > D	ΔD^M = -P	64	Household is in the process of repaying.
2.2	tB < year	B = D	ΔD^M = 0	16	No repayment yet. Interest only scheme or behind schedule and only able to pay interest rates.
2.3	tB < year	B < D	ΔD^M = -P	3	Household is behind on payments and interest accumulated.
				108	

tB is defined as X902 (X1002), B corresponds to X904 (X1004), D is X905 (X1005) and P=X908-(X916/10000*X905) and P=X1008-(X1016/10000*X1005). Terms in brackets refer to the third mortgage.

Information on the second ($D_{i,t}^{M2}$) and third mortgage ($D_{i,t}^{M3}$) is not as comprehensive as for the first mortgage ($D_{i,t}^{M1}$) which is discussed in section 2. In particular whether the mortgage was refinanced or not is not asked for second and third mortgages and thus a different way of defining the change in the outstanding liability is applied. The key difference is the assumption that all second and third mortgages taken out in the year of the interview are new debt and are not refinanced. The definition of cases for second and third mortgages taken out prior to the year of the interview is equivalent to the categorisation of the first mortgage. Cases are summarized in Table B1 which also provides the exact codes of the variables used. D and ΔD stand for $D_{i,t}^{M2}$ or $D_{i,t}^{M3}$ and $\Delta D_{i,t}^{M2}$ or $\Delta D_{i,t}^{M3}$ respectively. The reported number of observations refers to the second mortgage in implicate 3 of the 2004 wave.

6.2 Mortgages on other residential property ($D_{i,t}^{OM}$)

'Mortgages on other residential properties' refer to mortgages on vacation homes and other property which is used for residential purposes besides the main residence. Property which is not used as residential property is treated as a business asset and any liabilities on such assets are netted out with the current value. The SCF introduces this convention and it is also used for the purposes of this paper. Since information on how the money was used is not available, the year of the purchase of the property (tp) is used as the next best option. It is assumed that mortgages taken out in the current year represent new debt if the property was bought in the same year, otherwise they are refinanced. The justification for this assumption is that on the one hand, it only affects a small number of observations and on the other hand, households with other residential properties are wealthier and less likely to use these additional properties in order to secure borrowing. Using the year of the purchase as additional information adds one more step to the analysis. The SCF asks about up to three mortgages on residential property ($D_{i,t}^{OM} = D_{i,t}^{OM1} + D_{i,t}^{OM2} + D_{i,t}^{OM3}$). All cases for all three mortgages of this category are summarized in Table B2. D and ΔD stand for $D_{i,t}^{OM1}$, $D_{i,t}^{OM2}$ or $D_{i,t}^{OM3}$ and $\Delta D_{i,t}^{OM1}$, $\Delta D_{i,t}^{OM2}$ or $\Delta D_{i,t}^{OM3}$ respectively. The number of observations refers to $D_{i,t}^{OM1}$ in implicate 3 in 2004.

Table III-2: Changes in the outstanding amount of mortgages on other residential property

line	step 1	step 2	step 3	definition	N	interpretation
1	tB = year	tp=year		ΔD = D	45	Household used mortgage to buy property thus new debt.
2	tB = year	tp<year	B > D	ΔD = D-B	15	Assumption: mortgage was refinanced and already some repayment occurred.
3	tB = year	tp<year	B = D	ΔD = 0	12	Assumption: mortgage was refinanced but no repayment yet.
4	tB = year	tp<year	B < D	ΔD = D-B	0	Irrelevant. No observations fall into this case in any year.
5.1	tB <year		B > D	ΔD = -P	206	Household paying back.
5.2	tB < year		B = D	ΔD = 0	35	Household not paying back yet (due to interest only period).
5.3	tB < year		B < D	ΔD = -P	2	Household behind on schedule and interest rate accumulates.
					315	

tp indicates year property was bought. tB is defined as X1713 (X1813/X1913), B corresponds to X1714 (X1814/X1914), D is X1715 (X1815/X1915) and P=X1718-(X1726/10000*X1715). All stocks and payments are multiplied by the ownership share (X1705/X1805/X1905). Terms in brackets refer to the second and third residential property beyond the main residence.

Cases 1 and 5.1 to 5.3 are identical to the definitions for second and third mortgages on the primary residence. Cases 2 to 3 are different in that the payment is not inferred from the reported payments and interest rates of the household but rather by the difference between the amount initially borrowed and currently outstanding. The reason is that it is easier to remember outstanding amounts and the amount initially borrowed compared to changing interest rates and payments. Accordingly if the amount initially borrowed is different from the amount currently outstanding the difference is defined as principal repayment. In case the current amount exceeds the amount initially borrowed (case 4), this would indicate that the household fell behind in payments within the first year of the mortgage. In 2004 not a single observation fell into this unlikely case.

6.3 Consumer loans

The SCF collects information about up to six unsecured consumer loans[55]. The change in the outstanding amount is derived identically to the cases of second and third mortgages. The reason is that for consumer loans are usually not refinanced. Also the way the SCF asks about them in the interview does not allow for the possibility of refinancing (`How much was borrowed or financed, not counting the finance charges?')[56]. So the assumption that consumer loans taken out in the year of the interview represent new debt seems uncontroversial. The definitions are summarized in Table B3, the number of observations corresponds to the third implicate in the 2004 wave.

Table III-3: Changes in the outstanding amount of unsecured consumer loans

case	step 1	step 2	definition	N	interpretation
1	tB = year		ΔD = D	110	Taking on new debt.
2.1	tB <year	B > D	ΔD = -P	160	Paying back an existing consumer loan.
2.2	tB < year	B = D	ΔD = 0	32	Interest only period or household struggles with payments.
2.3	tB < year	B < D	ΔD = -P	6	Household behind schedule and interest accumulated.
				308	

tB is defined as X2713, B corresponds to X2714, D is X2723 and P=X2718-(X2724/10000*X2723). If X2724 is not observed X2719 is used, both expressions are transformed to annual frequencies. tB for the next five consumer loans corresponds to X2730/X2813/X2830/X2913/X2930, B for the next five corresponds to X2731/X2814 etc.

[55] Consumer loans are based on item X2730 as long as the loan was not taken out for investment purposes in non-residential real estate (X2710==78).
[56] Refers to item X2714 in the SCF codebook.

6.4 Car and vehicle loans

Loans to purchase cars and other vehicles are categorised in the same way as consumer loans. Debt taken on in the current period is assumed to be new debt. Definitions and number of observations for each case in the 2004 wave (implicate 3) are displayed in Table B4.

Table III-4: Changes in the outstanding amount of car and vehicle loans

case	step 1	step 2	definition	N	interpretation
1	tB = year		ΔD = D	284	Taking out a car loan in the current year thus new debt.
2.1	tB <year	B > D	ΔD = -P	932	Paying back an earlier car loan.
2.2	tB < year	B = D	ΔD = 0	3	Not paying on the loan.
2.3	tB < year	B < D	ΔD = -P	0	Accumulating interest failed to pay.
				1219	

tB is defined as X2208, B corresponds to X2209, D is X2218 and P=X2213-(X2219/10000*X2209). If X2213 is not observed X7537 is used, both expressions are transformed to annual frequencies. tB for the next five car loans corresponds to X2308/X2408/X2509/X2609/X7157, B for the next five corresponds to X2309/X2409 etc.

6.5 Education loans

The SCF collects information on up to six education or student loans. While the changes in the outstanding amount are defined in the same way as for car and consumer loans, the number of observations in each case in 2004 (implicate 3) demonstrates the different way in which student loans are paid back. With most student loans principal as well as interest payments only start after graduation. This applies to federal as well as private student loans. Federal student loans may be subsidised or unsubsidised. Federal student loans normally have lower interest rates than private ones and in addition, interest payments are paid for by the government in the case of subsidised loans. For student loans it is therefore much more common to observe that the outstanding amount equals the amount initially borrowed even if borrowing occurred in a year prior to the interview. The reason for this is that interest on subsidised federal student loans does not accumulate but is paid for by the government. Students with unsubsidised federal or private loans are typically faced with an outstanding amount after graduation which exceeds the amount initially borrowed due to the accumulated interest. This is reflected by the high number of observations falling in case 2.3 compared to consumer or car loans.

Table III-5: Changes in the outstanding amount of education loans

case	step 1	step 2	definition	N	interpretation
1	tB = year		ΔD = D	53	Student taking out a student loan in the current year thus new debt.
2.1	tB < year	B > D	ΔD = -P	254	Student no longer enrolled and paying back the loan.
2.2	tB < year	B = D	ΔD = 0	125	Student still enrolled, no accumulation of interest rates and not started to pay back.
2.3	tB < year	B < D	ΔD = -P	60	Non subsidised or private loan, interest accumulated, now in process of repaying.
				492	

tB is defined as X7804, B corresponds to X7805, D is X7824 and P=X7815-(X7822/10000*X7824). If X7815 is not observed X7817 is used, both expressions are transformed to annual frequencies. tB for the next five education loans corresponds to X7827/X7850/X7904/X7927/X7950.

6.6 Other loans

Other loans represent loans taken out to buy the primary residence or to undertake home improvements. They are categorized as `other´ because they are not owed to a bank or a financial institution but to relatives or to the seller of the property. By distinction, this refers to loans which are part of informal arrangements. In 2004 (implicate 3) a very limited number of households reported such liabilities and their importance for the aggregate picture is minimal.

Table III-6: Changes in the outstanding amount of other loans

case	step 1	step 2	definition	N	interpretation
1	tB = year		ΔD = D	5	Taking out new debt.
2.1	tB <year	B > D	ΔD = -P	16	In the process of repaying.
2.2	tB < year	B = D	ΔD = 0	10	No repayments yet.
2.3	tB < year	B < D	ΔD = -P	0	Irrelevant.
				31	

tB is defined as X1034, B corresponds to X1035, D is X1044 and P=X1039-(X1045/10000*X1044). If X1039 is not observed X1040 is used, both expressions are transformed to annual frequencies. tB for home improvement loans is X1205.

Appendix IV

Table VI-1. Data definitions and sources

abbreviation	full variable name	unit	source
Y^D	Disposable real gross income, household sector (deflated using PC)	national currency, billion	AMECO
C	Private final consumption expenditure at 2005 prices	national currency, billion	AMECO
PC	Price deflator private final consumption expenditure (PCPH)	2005=1	AMECO
R	Real long-term interest rates, deflator GDP	%	AMECO and OECD (MEI)
OLD	Fraction of population aged 65 and older	%	AMECO
D	Total credit to the household sector (deflated using PC)	national currency, billion	BIS
TOP1	Top 1% income share of the SWIID	%	SWIID v4
GINI	Gini coefficient (pre tax and post transfer) of the Standardized World Income Inequality Database		SWIID v5
PP	Real property prices BIS (exact definitions vary across countries, deflated using PC)	2005=1	BIS
SP	Share price index (deflated using PC)	2005=1	IMF (International Financial Statistics) and OECD (MEI)
CRED	Fraser Index, Subcategory 5A Credit Regulation: percentage of privately held deposits (higher values higher percentage), interest rate controls (market rates and positive real rates result in higher values), private sector credit (higher values less gov borrowing)	index between [0,10]	Fraser Institute
FIN	Index of financial reforms measuring: credit controls, interest rate controls, entry barriers, state ownership in banking, capital account restrictions, supervision of the banking sector and securities market policy. Policies in each of these 7 areas are awarded a number of 0 to 3 where higher numbers represent liberal policies.	index between [1,21]	IMF (Abiad et al. 2008 - A New Database of Financial Reforms)

Table IV-2. Data summary statistics I

Var		Mean	Std. Dev.	Min	Max	unit	Observations		
D	overall	24,543	81,513	20	356,783	national	N	=	325
	between		78,577	42	284,233	currency,	n	=	13
	within		22,648	-125,565	97,093	billion	T	=	25
C	overall	20,973	67,374	43	294,344	national	N	=	325
	between		66,506	61	240,736	currency,	n	=	13
	within		12,152	-56,981	74,580	billion	T	=	25
Y^D	overall	24,725	79,125	48	321,185	national	N	=	325
	between		78,559	67	284,327	currency,	n	=	13
	within		11,528	-53,682	61,584	billion	T	=	25
PP	overall	0.75	0.26	0.40	1.61		N	=	325
	between		0.20	0.59	1.26	2005=1	n	=	13
	within		0.16	0.46	1.24		T	=	25
SP	overall	0.68	0.38	0.07	2.19		N	=	325
	between		0.18	0.39	1.06	2005=1	n	=	13
	within		0.34	0.07	2.35		T	=	25
PC	overall	0.80	0.18	0.23	1.07		N	=	325
	between		0.08	0.67	0.99	2005=1	n	=	13
	within		0.17	0.35	1.15		T	=	25
R	overall	0.04	0.02	-0.08	0.14		N	=	320
	between		0.01	0.03	0.05	%	n	=	13
	within		0.02	-0.08	0.14		T	=	25
OLD	overall	0.14	0.02	0.09	0.21		N	=	325
	between		0.02	0.11	0.17	%	n	=	13
	within		0.02	0.09	0.21		T	=	25
TOP1	overall	0.08	0.03	0.03	0.18		N	=	325
	between		0.03	0.05	0.13	%	n	=	13
	within		0.02	0.03	0.18		T	=	25
GINI	overall	0.44	0.05	0.29	0.54	Gini	N	=	325
	between		0.04	0.37	0.52	coefficient	n	=	13
	within		0.03	0.36	0.54		T	=	25
CRED	overall	8.8	1.1	5.0	10.0	index	N	=	325
	between		0.8	6.8	9.6	[0,10]	n	=	13
	within		0.7	7.0	11.0		T	=	25
FIN	overall	16.9	4.1	2.0	21.0	index	N	=	312
	between		2.1	13.8	20.5	[1,21]	n	=	13
	within		3.6	3.1	23.1		T	=	24

Table IV-3: Unit root tests, levels

	P	L	Z	transformation	deterministic part
d	0.929	0.992	0.988	level	trend+const
c	0.770	0.947	0.942	level	trend+const
y^D	0.623	0.840	0.864	level	trend+const
pp	0.458	0.726	0.774	level	trend+const
sp	0.009	0.004	0.003	level	trend+const
TOP1	0.066	0.202	0.239	level	trend+const
GINI	0.624	0.686	0.707	level	trend+const
OLD	0.132	0.783	0.740	level	trend+const
CRED	0.687	0.761	0.765	level	trend+const
FIN	0.014	0.056	0.114	level	trend+const
R	0.362	0.586	0.579	level	trend+const

Panel unit root tests (H_0: all series contain unit roots) based on Choi (2001) who uses the following labels: inverse chi-square test (P), inverse normal test (Z) and logit test (L). P-values from ADF tests with 3 lags are combined. Lower case letters indicate variables are transformed by taking natural logarithms.

Table IV-4: Unit root tests, first differenced series

	P	L	Z	transformation	deterministic part
d	0.683	0.506	0.511	first differenced	trend+const
c	0.140	0.045	0.036	first differenced	trend+const
y^D	0.006	0.008	0.013	first differenced	trend+const
pp	0.015	0.025	0.041	first differenced	trend+const
sp	0.056	0.020	0.015	first differenced	trend+const
TOP1	0.000	0.000	0.000	first differenced	trend+const
GINI	0.036	0.026	0.026	first differenced	trend+const
OLD	0.000	0.000	0.083	first differenced	trend+const
CRED	0.003	0.002	0.002	first differenced	trend+const
FIN	0.011	0.006	0.005	first differenced	trend+const
R	0.000	0.000	0.000	first differenced	trend+const

Panel unit root tests (H_0: all series contain unit roots) based on Choi (2001) who uses the following labels: inverse chi-square test (P), inverse normal test (Z) and logit test (L). P-values from ADF tests with 3 lags are combined. Lower case letters indicate variables are transformed by taking natural logarithms.

Table IV-5: Individual country results differenced debt equation

betas		y^D	Q	pp	sp	R	OLD	cred	const
AU	TOP1	-0.16	-1.89	0.28	0.06	-0.45	-50.55	0.12	0.14
AU	GINI	-0.13	0.13	0.32	0.07	0.09	-44.44	0.08	0.12
BE	TOP1	0.40	-0.70	0.69	0.06	0.07	7.00	-0.12	0.00
BE	GINI	0.31	0.83	0.63	0.04	0.00	2.85	-0.15	0.01
CA	TOP1	1.09	0.44	0.35	0.05	-0.24	0.17	-0.86	0.01
CA	GINI	1.18	0.52	0.36	0.05	-0.32	-0.05	-0.82	0.01
DE	TOP1	0.89	-1.09	-0.27	0.04	0.18	-2.32	0.01	0.01
DE	GINI	0.99	-0.60	-0.47	0.03	-0.04	-2.36	0.09	0.01
FI	TOP1	1.02	-0.81	0.32	-0.04	0.16	28.10	0.43	-0.02
FI	GINI	0.99	-0.05	0.32	-0.04	0.19	28.58	0.42	-0.02
FR	TOP1	0.00	1.23	0.26	0.03	0.14	-5.13	-0.22	0.04
FR	GINI	0.06	0.06	0.28	0.04	0.23	-4.69	-0.25	0.04
UK	TOP1	-0.54	-1.09	0.41	-0.10	-0.14	-1.01	0.12	0.07
UK	GINI	-0.32	0.04	0.40	-0.12	0.02	-0.85	0.06	0.06
IT	TOP1	0.18	20.36	0.40	0.02	0.19	-4.14	-0.08	0.04
IT	GINI	0.96	1.18	0.39	0.06	-1.27	-1.98	-0.20	0.04
JP	TOP1	0.90	2.53	0.45	0.02	-0.18	-6.20	0.07	0.05
JP	GINI	0.68	0.26	0.52	0.03	-0.14	-5.29	0.05	0.05
NL	TOP1	-0.07	9.07	0.92	0.01	-0.43	19.62	-0.50	0.01
NL	GINI	0.20	-1.26	0.74	-0.03	-0.18	17.12	-0.09	0.02
NO	TOP1	1.49	-1.59	0.33	0.02	-0.19	26.15	0.49	0.01
NO	GINI	0.21	0.30	0.36	0.01	0.05	19.18	0.41	0.04
SE	TOP1	0.14	-4.36	0.56	0.02	-0.08	18.60	-0.01	0.02
SE	GINI	0.27	-0.39	0.52	0.02	0.12	18.03	-0.04	0.01
US	TOP1	0.65	-0.41	0.29	0.01	-0.97	8.03	-0.25	0.03
US	GINI	0.58	0.02	0.27	0.01	-1.01	7.07	-0.26	0.03

Estimated coefficients from estimating equation (5) by OLS for individual countries. Green indicates statistical significance at 1% level, yellow at 5% and orange at 10%. Lower case letters indicate logarithmic transformation.

Table IV-6: Growth regressions, with lagged residential property prices

	(1) 1980-2005	(2) 1980-2005	(3) 1980-2007	(4) 1980-2007
Δy^D	0.539***	0.527***	0.529***	0.451***
	(0.11)	(0.11)	(0.12)	(0.13)
Δpp_{t-1}	0.307***	0.310***	0.315***	0.328***
	(0.04)	(0.04)	(0.03)	(0.04)
Δsp	0.052***	0.051***	0.054***	0.053***
	(0.01)	(0.01)	(0.01)	(0.01)
$\Delta TOP1$	-0.163		-0.531***	
	(0.22)		(0.11)	
$\Delta GINI$		0.019		-0.018
		(0.20)		(0.17)
ΔR	0.196**	0.210***	0.167***	0.218***
	(0.07)	(0.05)	(0.05)	(0.06)
ΔOLD	-1.89	-1.898	-2.463	-2.57
	(2.25)	(2.17)	(1.77)	(1.73)
Δfin	-0.013	-0.013		
	(0.04)	(0.04)		
$\Delta cred$			-0.053	-0.056
			(0.04)	(0.04)
$const$	0.032***	0.032***	0.033***	0.034***
	(0.00)	(0.00)	(0.00)	(0.00)
N	280	280	306	306
adj. R^2	0.37	0.37	0.39	0.38
F-stat	74	83	104	276

Fixed effects (FE) estimations, dependent variable: Δd_{it}, robust standard errors clustered at the country level in brackets. Stars indicate statistical significance: * p<0.1, ** p<0.05, *** p<0.01. Lower case letters indicate variables are transformed by taking natural logarithms.

Table IV-7: Individual country results differenced consumption equation

		y^D	Q	pp	sp	R	OLD	cred	const
AU	TOP1	0.23	0.03	0.08	0.00	-0.14	-0.99	0.00	0.03
AU	GINI	0.22	0.05	0.08	0.00	-0.16	-0.30	-0.01	0.03
BE	TOP1	0.35	0.79	0.00	0.00	0.12	2.11	-0.09	0.01
BE	GINI	0.36	-0.34	0.03	0.01	0.17	4.38	-0.05	0.00
CA	TOP1	0.56	-0.82	0.07	0.06	0.06	-1.92	-0.06	0.02
CA	GINI	0.44	-0.43	0.07	0.05	0.13	-1.78	-0.13	0.02
DE	TOP1	0.68	2.25	0.59	-0.03	0.19	-2.83	0.14	0.02
DE	GINI	0.44	2.04	1.24	0.02	1.10	-3.00	-0.13	0.03
FI	TOP1	0.29	0.47	0.18	0.00	0.00	4.53	0.01	0.01
FI	GINI	0.27	-0.18	0.17	0.00	0.02	4.45	0.00	0.01
FR	TOP1	0.53	1.56	0.02	0.01	0.01	-1.84	-0.01	0.01
FR	GINI	0.55	-0.01	0.04	0.01	0.07	-1.47	-0.04	0.01
UK	TOP1	0.51	0.87	0.16	-0.02	0.05	-0.29	-0.05	0.01
UK	GINI	0.31	0.51	0.18	0.00	-0.21	-0.40	0.02	0.01
IT	TOP1	0.53	5.08	0.06	0.02	0.10	0.95	-0.01	0.00
IT	GINI	0.65	0.14	0.05	0.03	-0.23	1.26	-0.03	0.00
JP	TOP1	0.51	0.94	0.07	0.00	0.04	-2.06	0.05	0.02
JP	GINI	0.42	-0.03	0.11	0.01	0.02	-1.53	0.04	0.02
NL	TOP1	0.06	0.56	0.19	0.03	-0.22	-7.00	-0.09	0.02
NL	GINI	0.20	-1.71	-0.07	0.01	-0.08	-10.91	0.23	0.03
NO	TOP1	0.46	-0.44	0.12	0.03	0.04	-5.22	-0.12	0.01
NO	GINI	0.11	0.22	0.13	0.03	0.11	-7.21	-0.13	0.01
SE	TOP1	0.02	-0.65	0.21	0.02	0.21	4.28	0.00	0.01
SE	GINI	-0.12	0.22	0.24	0.01	0.15	5.18	0.03	0.01
US	TOP1	0.61	0.21	0.03	0.04	-0.21	-1.22	-0.05	0.01
US	GINI	0.65	0.45	0.05	0.04	-0.21	-1.08	-0.04	0.01

Estimated coefficients from estimating equation (6) by OLS for individual countries. Green indicates statistical significance at 1% level, yellow at 5% and orange at 10%. Lower case letters indicate logarithmic transformation.

Appendix V

This appendix contains the Stata code used to create the datasets used in chapters 3 and 4.

```
************************************************
*Section 1: Combining public and summary dataset for DEMOGRAPHICS
*Section 2: Combining public and summary dataset for INCOME
*Section 3: Combining public and summary dataset for ASSETS
*Section 4: Combining public and summary dataset for DEBT
*Section 5: Combine everything into a general data set
*Section 6: Set up imputation structure
*Section 7: Creation of additional variables
*Section 8: Define Relative Income Measures
*Section 9: IHS transform the data

***********construct SCF dataset************
clear
*set directory to data folder
cd "C:\Users\Rafael\Google Drive\Data\SCF Data\original files"
*cd "H:\Data\SCF - Home Drive\original files" //in remote or Kingston desktop mode

*********program for converting regular payments
capture program drop interval
program define interval
                args amount interval
                gen `amount't=`amount'*365 if `interval'==1
                replace `amount't=`amount'*52 if `interval'==2
                replace `amount't=`amount'*26 if `interval'==3
                replace `amount't=`amount'*12 if `interval'==4
                replace `amount't=`amount'*4 if `interval'==5
                replace `amount't=`amount'*1 if `interval'==6
                replace `amount't=`amount'*2 if `interval'==11
                replace `amount't=`amount'*6 if `interval'==12
                replace `amount't=`amount'/3 if `interval'==13
                replace `amount't=`amount'*24 if `interval'==31
                replace `amount't=`amount'*13 if `interval'==23
                replace `amount't=`amount'*52/6 if `interval'==24
```

```
                replace `amount't=`amount'*0 if `interval'==0 | `interval'==-7 | `interval'==-2 | `interval'==-1 | `interval'==22 | `interval'==8
end


*****************************************************************************************
*****Section 1: Combining public and summary dataset for DEMOGRAPHICS*****************
*****************************************************************************************

*NOTE: This does not only deal with demographics but also merges the replicate weights
*          into the data set. (And thus yields large files!).

*****load 1989 dataset
clear
use "p1989i6.dta"
*X6536: using documents when answering, not part of 1989 sample
*X6770: How many years lived within 25mile of current home, not part of 1989 sample
*X6780 X6784 X6781 X6785: unemployment of head and partner, not part of 1989 sample
merge 1:1 X1 using "rscfp1989.dta"
gen year=1989
keep X1 XX1 ///
X301 ///
X4511 X5111 ///
X101 ///
age hhsex lf educ edcl race racecl kids married OCCAT1 OCCAT2 indcat year housecl wgt
merge m:1 XX1 using "p1989_rw1.dta"
forvalues i=1/999 {
gen repwgt`i'=WT1B`i'*MM`i'
}
drop WT1B* MM*
save "combined1989demo.dta" , replace

*****load 1992 dataset
clear
use "p1992i6.dta"
*repeat everything for the 1992 dataset*************************************************
merge 1:1 Y1 using "rscfp1992.dta"
*X6770 X6780 X6784 X6781 X6785, not part of 1992 sample
gen year=1992
```

```
keep YY1 Y1 X6536 X30022 X30074 ///
X301 ///
X4511 X5111 ///
X101 ///
age hhsex lf educ edcl race racecl kids married OCCAT1 OCCAT2 indcat year housecl wgt
merge m:1 YY1 using "p1992_rw1.dta"
forvalues i=1/999 {
gen repwgt`i'=WT1B`i'*MM`i'
}
drop WT1B* MM*
save "combined1992demo.dta" , replace

*****load 1995 dataset
clear
use "p1995i6.dta"
*repeat everything for the 1995 dataset************************************************
merge 1:1 Y1 using "rscfp1995.dta"
*X6770 X6780 X6784 X6781 X6785
gen year=1995
keep YY1 Y1 X6536 X30022 X30074 ///
X301 ///
X4511 X5111 ///
X101 ///
X7000 X8000 X7002 ///
age hhsex lf educ edcl race racecl kids married OCCAT1 OCCAT2 indcat year housecl wgt
merge m:1 YY1 using "p1995_rw1.dta"
forvalues i=1/999 {
gen repwgt`i'=WT1B`i'*MM`i'
}
drop WT1B* MM*
save "combined1995demo.dta" , replace

*****load other datasets
*year 1998 needs to be extra cause last year which reports X30022 X30074
clear
use "p1998i6.dta"
merge 1:1 Y1 using "rscfp1998.dta"
```

```
gen year=1998
keep YY1 Y1 X6536 X30022 X30074 ///
X301 ///
X6770 ///
X6780 X6784 X6781 X6785 ///
X4511 X5111 ///
X101 ///
X7000 X8000 X7002 X7017 ///
age hhsex lf educ edcl race racecl kids married OCCAT1 OCCAT2 indcat year housecl wgt
merge m:1 YY1 using "p1998_rw1.dta"
	forvalues i=1/999 {
		gen repwgt`i'=WT1B`i'*MM`i'
	}
drop WT1B* MM*
save "combined1998demo.dta" , replace
*year 2001 needs to be extra because in the replicate dataset all variables are in lower case letters
clear
use "p2001i6.dta"
merge 1:1 Y1 using "rscfp2001.dta"
gen year=2001
keep YY1 Y1 X6536 ///
X301 ///
X6770 ///
X6780 X6784 X6781 X6785 ///
X4511 X5111 ///
X101 ///
X7000 X8000 X7002 X7017 ///
age hhsex lf educ edcl race racecl kids married OCCAT1 OCCAT2 indcat year housecl wgt
merge m:1 YY1 using "p2001_rw1.dta"
forvalues i=1/999 {
	gen repwgt`i'=wt1b`i'*mm`i'
}
drop wt1b* mm*
save "combined2001demo.dta" , replace
*all other years
foreach year in 2004 2007 2010 2013 {
	clear
	use "p`year'i6.dta"
```

```
		merge 1:1 Y1 using "rscfp`year'.dta"
		gen year=`year'
		keep YY1 Y1 X6536 ///
		X301 ///
		X6770 ///
		X6780 X6784 X6781 X6785 ///
		X4511 X5111 ///
		X101 ///
		X7000 X8000 X7002 X7017 ///
		age hhsex lf educ edcl race racecl kids married OCCAT1 OCCAT2 indcat year housecl wgt
		merge m:1 YY1 using "p`year'_rw1.dta"
		forvalues i=1/999 {
				gen repwgt`i'=WT1B`i'*MM`i'
		}
		drop WT1B* MM*
		save "combined`year'demo.dta" , replace
}
*
********Section 1.2: combining the different waves into repeated cross section***********
*this part needs to be run on 64 bit stata
clear
use "combined1989demo.dta"
append using "combined1992demo.dta"
append using "combined1995demo.dta"
append using "combined1998demo.dta"
append using "combined2001demo.dta"
append using "combined2004demo.dta"
append using "combined2007demo.dta"
append using "combined2010demo.dta"
append using "combined2013demo.dta"
drop _merge

*check missing variables
*mdesc
*575 observations for X6536 (die R use documents in answering the questions) are missing in 1992

*strangely in the 1989 sample the numbering of their X1 variables jumps from 2324 to 3001 which
*is in stata at the 11620th observation.
```

```
*Nevertheless I will combine X1 with Y1
replace Y1=X1 if Y1==.
replace YY1=XX1 if YY1==.
drop X1 XX1
*mdesc
gen imp=Y1-10*YY1
saveold "combined_allyears_demo.dta" , replace
foreach year in 1989 1992 1995 1998 2001 2004 2007 2010 2013 {
		erase "combined`year'demo.dta"
}
*

****************************************************************************************
********Section 2: Combining public and summary dataset for INCOME*********************
****************************************************************************************

*****load 1989 dataset
clear
use "p1989i6.dta"
*repeat everything for the 1989 dataset*************************************************
merge 1:1 X1 using "rscfp1989.dta"
*not in the 1989 wave: X7650 X7509 X7507 wsaved saved spendmor
gen year=1989
keep X1 XX1 X304 late LATE60 X3004 X3005 X407 tpay mortpay conspay revpay pirtotal pirmort pircons pirrev PIR40 ///
income kginc wageinc bussefarminc intdivinc norminc ///
penacctwd ssretinc transfothinc year
save "combined1989income.dta" , replace

*****load 1992 dataset
clear
use "p1992i6.dta"
*repeat everything for the 1992 dataset*************************************************
merge 1:1 Y1 using "rscfp1992.dta"
gen year=1992
*not in 1992 wave: spendmor
keep YY1 Y1 X7650 X304 ///
X7509 X7507 tpay mortpay conspay revpay pirtotal pirmort pircons pirrev PIR40 ///
income kginc wageinc bussefarminc intdivinc norminc late LATE60 X3004 X3005 X407 ///
```

```
wsaved saved  penacctwd ssretinc transfothinc year
save "combined1992income.dta" , replace

*****load 1995 dataset
clear
use "p1995i6.dta"
*repeat everything for the 1995 dataset**************************************************
merge 1:1 Y1 using "rscfp1995.dta"
*not in 1995 wave: spendmor
gen year=1995
keep YY1 Y1 X7650 X304 ///
X7509 X7507 tpay mortpay conspay revpay pirtotal pirmort pircons pirrev PIR40 ///
income kginc wageinc bussefarminc intdivinc norminc late LATE60 X3004 X3005 X407 ///
wsaved saved penacctwd ssretinc transfothinc year
save "combined1995income.dta" , replace

*now run loops for all years
foreach year in 1998 2001 2004 2007 2010 2013 {
	clear
	use "p`year'i6.dta"
	keep YY1 Y1 X7650 X7509 X7507 X304 X3004 X3005 X407
	merge 1:1 Y1 using "rscfp`year'.dta"
	keep YY1 Y1 X7650 X304 ///
	X7509 X7507 late LATE60 X3004 X3005 X407 tpay mortpay conspay revpay pirtotal pirmort pircons pirrev PIR40 ///
	income kginc wageinc bussefarminc intdivinc norminc ///
	wsaved saved spendmor penacctwd ssretinc transfothinc
	gen year=`year'
	save "combined`year'income.dta" , replace
}
*

********Section 2.2: combining the different waves into repeated cross section***********
clear
use "combined1989income.dta"
append using "combined1992income.dta"
append using "combined1995income.dta"
append using "combined1998income.dta"
append using "combined2001income.dta"
```

```
append using "combined2004income.dta"
append using "combined2007income.dta"
append using "combined2010income.dta"
append using "combined2013income.dta"

*strangely in the 1989 sample the numbering of their X1 variables jumps from 2324 to 3001 which
*is in stata at the 11620th observation.
*Nevertheless I will combine X1 with Y1
replace Y1=X1 if Y1==.
replace YY1=XX1 if YY1==.
drop X1 XX1
gen imp=Y1-10*YY1
mdesc
save "combined_allyears_income.dta" , replace
foreach year in 1989 1992 1995 1998 2001 2004 2007 2010 2013 {
		erase "combined`year'income.dta"
}
*

***************************************************************************************
*****Section 3: Combining public and summary dataset for ASSETS******************
***************************************************************************************

*****load 1989 dataset
clear
use "p1989i6.dta"
merge 1:1 X1 using "rscfp1989.dta"
keep X1 XX1 X720 X717 X718 X716 ///
X604 X607 X606 X608 ///
X614 X617 X616 X618 ///
X627 X626 X628 ///
X631 X630 X632 X634 ///
X1703 X1705 X1706 X1708 X1709 X1710 ///
X1803 X1805 X1806 X1808 X1809 X1810 ///
X1903 X1905 X1906 X1908 X1909 X1910 ///
X3110 X3210 X3310 ///
X5805 X5810 X5815 ///
X5804 X5809 X5814 ///
```

```
liq cds nmmf stocks bond retqliq savbnd cashli othma othfin equitinc fin ///
vehic houses oresre housecl nnresre bus othnfin nfin asset ///
kghouse kgore kgbus FARMBUS_KG kgstmf kgtotal
gen year=1989
gen cpi=0.5532286213
*creating variabls for current year bought properties
qui forvalues year=1989(3)1989 {
	gen REP_houses= (X607*(X606==`year')*(X608==5)+ /// X607 cost when achquired, X606 year of purchase,X608 purch=5
		X617*(X616==`year')*(X618==5)+ ///
		X627*(X626==`year')*(X628==5)+ ///
		X631*(X630==`year')*(X632==5)+ ///
		X717*(X720==`year')*(X718==5))/cpi
	gen capcor_houses=((X604-X607)*(X606==`year')*(X608==5)+ ///this variable corrects HW for capital gains losses which occured since the purchase (if that was in the current period)
		(X614-X617)*(X616==`year')*(X618==5)+ ///X627 and X631 can't be dealt with cause home and site not reported separately
		(X716-X717)*(X720==`year')*(X718==5))/cpi

	gen REP_oresre= ((X1703==12 | X1703==21 | X1703==40 | X1703==41 | X1703==42 | X1703==49 | X1703==50 | X1703==999)*(X1705/10000)*(X1708==`year')*(X1710==5)*X1709+ /// X1709
value when purchased
		(X1803==12 | X1803==21 | X1803==40 | X1803==41 | X1803==42 | X1803==49 | X1803==50 | X1803==999)*(X1805/10000)*(X1808==`year')*(X1810==5)*X1809+ ///
		(X1903==12 | X1903==21 | X1903==40 | X1903==41 | X1903==42 | X1903==49 | X1903==50 | X1903==999)*(X1905/10000)*(X1908==`year')*(X1910==5)*X1909)/cpi
	gen capcor_oresre=((X1703==12 | X1703==21 | X1703==40 | X1703==41 | X1703==42 | X1703==49 | X1703==50 | X1703==999)*(X1705/10000)*(X1708==`year')*(X1710==5)*(X1706-X1709)+ ///
X1706 current value
		(X1803==12 | X1803==21 | X1803==40 | X1803==41 | X1803==42 | X1803==49 | X1803==50 | X1803==999)*(X1805/10000)*(X1808==`year')*(X1810==5)*(X1806-X1809)+ ///
		(X1903==12 | X1903==21 | X1903==40 | X1903==41 | X1903==42 | X1903==49 | X1903==50 | X1903==999)*(X1905/10000)*(X1908==`year')*(X1910==5)*(X1906-X1909))/cpi
	gen REP=REP_houses+REP_oresre
	noi sum REP
	noi sum REP if REP<10,d
	gen HW=houses+oresre-REP_houses-REP_oresre-capcor_houses-capcor_oresre
	noi sum HW if HW<10,d
	replace HW=0 if HW<10
}
save "combined1989assets.dta" , replace

*load all the others
clear
foreach year in 1992 1995 1998 2001 2004 2007 2010 2013 {
	noi disp `year'
	clear
```

```
use "p`year'i6.dta"
merge 1:1 Y1 using "rscfp`year'.dta"
if `year'<2010 {
		keep Y1 YY1 ///
		X720 X717 X718 X716 ///
		X604 X607 X606 X608 ///
		X614 X617 X616 X618 ///
		X627 X626 X628 ///
		X631 X630 X632 X634 ///
		X1703 X1705 X1706 X1708 X1709 X1710 ///
		X1803 X1805 X1806 X1808 X1809 X1810 ///
		X1903 X1905 X1906 X1908 X1909 X1910 ///
		X3110 X3210 X3310 ///
		X5805 X5810 X5815 X5804 X5809 X5814 ///
		liq cds nmmf stocks bond retqliq savbnd cashli othma othfin equitinc fin ///
		vehic houses oresre housecl nnresre bus othnfin nfin asset ///
		kghouse kgore kgbus FARMBUS_KG kgstmf kgtotal
		}
else if `year'>=2010 {
		keep Y1 YY1 X42001 ///
		X720 X717 X718 X716 ///
		X604 X607 X606 X608 ///
		X614 X617 X616 X618 ///
		X627 X626 X628 ///
		X631 X630 X632 X634 ///
		X1703 X1705 X1706 X1708 X1709 X1710 ///
		X1803 X1805 X1806 X1808 X1809 X1810 ///
		X3110 X3210 ///
		X5805 X5810 X5815 X5804 X5809 X5814 ///
		liq cds nmmf stocks bond retqliq savbnd cashli othma othfin equitinc fin ///
		vehic houses oresre housecl nnresre bus othnfin nfin asset ///
		kghouse kgore kgbus FARMBUS_KG kgstmf kgtotal
}
gen year=`year'
gen cpi=.
if `year'==1992 replace cpi=0.6154741129
else if `year'==1995 replace cpi=0.6588132635
else if `year'==1998 replace cpi=0.6995346131
```

```
	else if `year'==2001 replace cpi=0.7614892379
	else if `year'==2004 replace cpi=0.810936591
	else if `year'==2007 replace cpi=0.8906340896
	else if `year'==2010 replace cpi=0.9331006399
	else if `year'==2013 replace cpi=1
	*creating dummies and variabls for last year bought properties
	gen REP_houses= (X607*(X606==`year')*(X608==5)+ /// X607 cost when achquired, X606 year of purchase,X608 purch=5
				X617*(X616==`year')*(X618==5)+ ///
				X627*(X626==`year')*(X628==5)+ ///
				X631*(X630==`year')*(X632==5)+ ///
				X717*(X720==`year')*(X718==5))/cpi
	gen capcor_houses=((X604-X607)*(X606==`year')*(X608==5)+ ///this variable corrects HW for capital gains losses which occured since the purchase (if that was in the current period)
				(X614-X617)*(X616==`year')*(X618==5)+ ///X627 and X631 can't be dealt with cause home and site not reported separately
				(X716-X717)*(X720==`year')*(X718==5))/cpi

	if `year'<=2007 {
		gen REP_oresre= ((X1703==12 | X1703==21 | X1703==40 | X1703==41 | X1703==42 | X1703==49 | X1703==50 | X1703==999)*(X1705/10000)*(X1708==`year')*(X1710==5)*X1709+ ///
				(X1803==12 | X1803==21 | X1803==40 | X1803==41 | X1803==42 | X1803==49 | X1803==50 | X1803==999)*(X1805/10000)*(X1808==`year')*(X1810==5)*X1809+ ///
				(X1903==12 | X1903==21 | X1903==40 | X1903==41 | X1903==42 | X1903==49 | X1903==50 | X1903==999)*(X1905/10000)*(X1908==`year')*(X1910==5)*X1909)/cpi
		gen capcor_oresre=((X1703==12 | X1703==21 | X1703==40 | X1703==41 | X1703==42 | X1703==49 | X1703==50 | X1703==999)*(X1705/10000)*(X1708==`year')*(X1710==5)*(X1706-
X1709)+ ///
				(X1803==12 | X1803==21 | X1803==40 | X1803==41 | X1803==42 | X1803==49 | X1803==50 | X1803==999)*(X1805/10000)*(X1808==`year')*(X1810==5)*(X1806-X1809)+ ///
				(X1903==12 | X1903==21 | X1903==40 | X1903==41 | X1903==42 | X1903==49 | X1903==50 | X1903==999)*(X1905/10000)*(X1908==`year')*(X1910==5)*(X1906-X1909))/cpi
	}
	else if `year'>=2010 {
		gen REP_oresre= ((X1703==12 | X1703==21 | X1703==40 | X1703==41 | X1703==42 | X1703==49 | X1703==50 | X1703==999)*(X1705/10000)*(X1708==`year')*(X1710==5)*X1709+ ///
				(X1803==12 | X1803==21 | X1803==40 | X1803==41 | X1803==42 | X1803==49 | X1803==50 | X1803==999)*(X1805/10000)*(X1808==`year')*(X1810==5)*X1809)/cpi
		gen capcor_oresre= ((X1703==12 | X1703==21 | X1703==40 | X1703==41 | X1703==42 | X1703==49 | X1703==50 | X1703==999)*(X1705/10000)*(X1708==`year')*(X1710==5)*(X1706-
X1709)+ ///
				(X1803==12 | X1803==21 | X1803==40 | X1803==41 | X1803==42 | X1803==49 | X1803==50 | X1803==999)*(X1805/10000)*(X1808==`year')*(X1810==5)*(X1806-X1809))/cpi
	}
	gen REP=REP_houses+REP_oresre
	gen HW=houses+oresre-REP_houses-REP_oresre-capcor_houses-capcor_oresre
	noi sum HW if HW<10,d
	replace HW=0 if HW<10
	*Basically in all cases where HW<0 the negative values are only very very small and only due to rounding errors.
	*Similar rounding errors occur for HW>0 but with very low values. Also these are set to HW=0
	save "combined`year'assets.dta" , replace
```

```
}
********Section 3.2: combining the different waves into repeated cross section***********
clear
use "combined1989assets.dta"
append using "combined1992assets.dta"
append using "combined1995assets.dta"
append using "combined1998assets.dta"
append using "combined2001assets.dta"
append using "combined2004assets.dta"
append using "combined2007assets.dta"
append using "combined2010assets.dta"
append using "combined2013assets.dta"
drop X42001
*adjusting for inflation
replace X717=X717/cpi
replace X716=X716/cpi
replace X607=X607/cpi
replace X604=X604/cpi
replace X617=X617/cpi
replace X614=X614/cpi
replace X627=X627/cpi
replace X631=X631/cpi
replace X1706=X1706/cpi
replace X1709=X1709/cpi
replace X1806=X1806/cpi
replace X1809=X1809/cpi
replace X1906=X1906/cpi
replace X1909=X1909/cpi
replace X5804=X5804/cpi
replace X5809=X5809/cpi
replace X5814=X5814/cpi
mdesc
*strangely in the 1989 sample the numbering of their X1 variables jumps from 2324 to 3001 which
*is in stata at the 11620th observation.
*Nevertheless I will combine X1 with Y1
replace Y1=X1 if Y1==.
replace YY1=XX1 if YY1==.
drop X1 XX1
```

```
gen imp=Y1-10*YY1
mdesc
save "combined_allyears_assets.dta", replace
*foreach year in 1989 1992 1995 1998 2001 2004 2007 2010 2013 {
*erase "combined`year'assets.dta"
*}
*




*************************************************************************************
*****Section 4: Combining public and summary dataset for DEBT******************
*************************************************************************************

***all years
quie foreach year of numlist 1989(3)2013 {
	noi disp .
	noi disp .
	noi disp `year'
	noi disp .
	noi disp .
	cd "C:\Users\Rafael\Google Drive\Data\SCF Data\original files"
	clear
	use "p`year'i6.dta"
	if `year'==1989 {
		merge 1:1 X1 using "rscfp`year'.dta"
		gen Y1=X1
		gen YY1=XX1
	}
	else {
		merge 1:1 Y1 using "rscfp`year'.dta"
	}
*********************************************************compute change in debt for primary mortgage X805
	noi disp "		primary mortgage"
	if `year'>=1995 & `year'<=2001 {

		gen tB=X802
```

```
	gen B=X804
	gen D=X805
	gen regu=X808
	gen freregu=X809
	gen typi=X813
	gen fretypi=X814
	gen inte=X816/10000
	quietly interval regu freregu
	quietly interval typi fretypi
	gen pD=regut*(regut>0)+typit*(typit>0)
	gen rD=D*inte*(inte>0)
	gen use=X7137

	*case 1: change in debt = D-B
	gen dmor802=D-B if tB==`year' & use==1 & D>0 & D!=.
	*case2: change in debt = extr
	replace dmor802=. if tB==`year' & (use==2 | use==3) & D>0 & D!=.
	*case3: change in debt = D
	replace dmor802=D if tB==`year' & (use==4 | use==8 | use==0) & D>0 & D!=.
	*case 4: change in debt = rD-pD
	replace dmor802=rD-pD if  tB<`year' & (B>D | B<D) & D>0 & D!=.
	*case 5: change in debt = 0
	replace dmor802=0 if tB<`year' & B==D & D>0 & D!=.
	*all which do not have a mortgage
	replace dmor802=0 if D==0

	nois count if dmor802==.
	drop tB B D regu regut freregu typi typit fretypi inte pD rD use
	rename dmor802 dmor1
}
if `year'>2001 {

	gen tB=X802
	gen B=X804
	gen D=X805
	gen regu=X808
	gen freregu=X809
	gen typi=X813
```

```
		gen fretypi=X814
		gen inte=X816/10000
		quietly interval regu freregu
		quietly interval typi fretypi
		gen pD=regut*(regut>0)+typit*(typit>0)
		gen rD=D*inte*(inte>0)
		gen use=X7137
		gen ex=X7138

		*case 1: change in debt = D-B
		gen dmor802=D-B if tB==`year' & use==1 & D>0 & D!=.
		*case2: change in debt = extr
		replace dmor802=ex if tB==`year' & use==2 & D>0 & D!=.
		*case2a: change in debt = extr+D-B
		replace dmor802=ex+D-B if tB==`year' & use==3 & D>0 & D!=.
		*case3: change in debt = D
		replace dmor802=D if tB==`year' & (use==4 | use==8 | use==0) & D>0 & D!=.
		*case 4: change in debt = rD-pD
		replace dmor802=rD-pD if  tB<`year' & (B>D | B<D) & D>0 & D!=.
		*case 5: change in debt = 0
		replace dmor802=0 if tB<`year' & B==D & D>0 & D!=.
		*all which do not have a mortgage
		replace dmor802=0 if D==0

		noi count if dmor802==.
		drop tB B D regu regut freregu typi typit fretypi inte pD rD use ex
		rename dmor802 dmor1
	}
***********************************************compute change for second and third mortgages X905/X1005
	noi disp "		second and third mortgage"
	if `year'>1992 {
		foreach var in 902 1002 {

			local aux2=`var'+2
			local aux3=`var'+3
			local aux6=`var'+6
			local aux7=`var'+7
			local aux11=`var'+11
```

```
			local aux12=`var'+12
			local aux14=`var'+14

			gen tB=X`var'
			gen B=X`aux2'
			gen D=X`aux3'
			gen inte=X`aux14'/10000
			gen regu=X`aux6'
			gen freregu=X`aux7'
			gen typi=X`aux11'
			gen fretypi=X`aux12'
			quietly interval regu freregu
			quietly interval typi fretypi
			gen pD=regut*(regut>0)+typit*(typit>0)
			gen rD=D*inte*(inte>0)
			gen tpurch=X720

			disp `year'
			*case1: change in debt = D
			gen dmor`var'=D if tB==`year' & D>0 & D!=.
			*case2: change in debt = rD-pD
			replace dmor`var'=rD-pD if tB<`year' & (B>D | B<D) & D>0 & D!=.
			*case3: change in debt = 0
			replace dmor`var'=0 if tB<`year' & B==D & D>0 & D!=.
			*all which do not have a mortgage
			replace dmor`var'=0 if D==0

			noi count if dmor`var'==.
			drop tB B D pD rD inte regu regut freregu typi typit fretypi tpurch
		}
		gen dmor=dmor1+dmor902+dmor1002
	}
********************************************************************compute change for other mortgages
	noi disp "		mortgages other residential properties"
	if `year'>1992 {
		foreach var in 1705 1805 {
			local aux1=`var'+10
			local aux2=`var'+9
```

```
local aux3=`var'+8
local aux4=`var'+13
local aux5=`var'+18
local aux6=`var'+21
local aux7=`var'+14
local aux8=`var'+19
local aux9=`var'-2
local aux10=`var'+3

*these commands convert the regular and typical payments into annual values
quietly interval X`aux4' X`aux7'
quietly interval X`aux5' X`aux8'

*pD is the principal plus interest repaid based on the answers of the respondent (the regular or typical payments)
*rD is the interest payment based on the interest rate and current outstanding amount
*resdum is a dummy equal to 1 if the underlying debt is not netted out with nnresre (netted out if non residential use)
gen pD=(X`aux4't*(X`aux4't>0)+X`aux5't*(X`aux5't>0))*(X`var'/10000)
gen rD=X`aux1'*(X`var'/10000)*(X`aux6'*(X`aux6'>0)/10000)
gen resdum=1*(X`aux9'==12|X`aux9'==21|X`aux9'==40|X`aux9'==41|X`aux9'==42|X`aux9'==49|X`aux9'==50|X`aux9'==53|X`aux9'==999)
gen B=X`aux2'*(X`var'/10000)
gen D=X`aux1'*(X`var'/10000)
gen tB=X`aux3'
gen tpurch=X`aux10'

*case1: change in debt = D
gen domor`var'=D if resdum==1 & tB==`year' & tpurch==`year' & D>0 & D!=.
*case2: change in debt = D-B
replace domor`var'=D-B if resdum==1 & tB==`year' & tpurch<`year' & (B>D | B<D) & D>0 & D!=.
*case3: change in debt = 0
replace domor`var'=0 if resdum==1 & tB==`year' & tpurch<`year' & B==D & D>0 & D!=.
*case4: change in debt = rD-pD
replace domor`var'=rD-pD if resdum==1 & tB<`year' & (B>D | B<D) & D>0 & D!=.
*case5: change in debt = 0
replace domor`var'=0 if resdum==1 & tB<`year' & B==D & D>0 & D!=.
*all which do not have a mortgage
replace domor`var'=0 if D==0
replace domor`var'=0 if resdum==0 & D>0 & D!=.
```

```
		noi count if domor`var'==.
		*the reason for the discrepancy here is that other mortgages are netted out with assets if resdum!=1
		drop pD rD resdum B D tB tpurch
}
if `year'<=2007 {
		foreach var in 1905 {
				local aux1=`var'+10
				local aux2=`var'+9
				local aux3=`var'+8
				local aux4=`var'+13
				local aux5=`var'+18
				local aux6=`var'+21
				local aux7=`var'+14
				local aux8=`var'+19
				local aux9=`var'-2
				local aux10=`var'+3

				quietly interval X`aux4' X`aux7'
				quietly interval X`aux5' X`aux8'

				gen pD=(X`aux4't*(X`aux4't>0)+X`aux5't*(X`aux5't>0))*(X`var'/10000)
				gen rD=X`aux1'*(X`var'/10000)*(X`aux6'*(X`aux6'>0)/10000)
				gen resdum=1*(X`aux9'==12|X`aux9'==21|X`aux9'==40|X`aux9'==41|X`aux9'==42|X`aux9'==49|X`aux9'==50|X`aux9'==53|X`aux9'==999)
				gen B=X`aux2'*(X`var'/10000)
				gen D=X`aux1'*(X`var'/10000)
				gen tB=X`aux3'
				gen tpurch=X`aux10'

				*case1: change in debt = D
				gen domor`var'=D if resdum==1 & tB==`year' & tpurch==`year' & D>0 & D!=.
				*case2: change in debt = D-B
				replace domor`var'=D-B if resdum==1 & tB==`year' & tpurch<`year' & (B>D | B<D) & D>0 & D!=.
				*case3: change in debt = 0
				replace domor`var'=0 if resdum==1 & tB==`year' & tpurch<`year' & B==D & D>0 & D!=.
				*case4: change in debt = rD-pD
				replace domor`var'=rD-pD if resdum==1 & tB<`year' & (B>D | B<D) & D>0 & D!=.
				*case5: change in debt = 0
				replace domor`var'=0 if resdum==1 & tB<`year' & B==D & D>0 & D!=.
```

```
					*all which do not have a mortgage
					replace domor`var'=0 if D==0
					replace domor`var'=0 if resdum==0 & D>0 & D!=.

					noi count if domor`var'==.
					drop pD rD resdum B D tB tpurch

				}
				gen domor=domor1705+domor1805+domor1905
			}
			else if `year'>=2010 {
				gen domor=domor1705+domor1805
			}
		}
**********************************************************************compute changes for consumer loans
		noi disp "		consumer loans"
		if `year'>1992 {
			gen XX2711=X7527
			gen XX2728=X7526
			gen XX2811=X7525
			gen XX2828=X7524
			gen XX2911=X7523
			gen XX2928=X7522

			foreach var in  2710 2727 2810 2827 2910 2927 {

				local aux1=`var'+1
				local aux3=`var'+3
				local aux4=`var'+4
				local aux8=`var'+8
				local aux9=`var'+9
				local aux10=`var'+10
				local aux13=`var'+13
				local aux14=`var'+14

				gen B=X`aux4'
				gen D=X`aux13'
				gen tB=X`aux3'
```

```
			gen inte=X`aux14'/10000
			gen regu=X`aux8'
			gen freregu=XX`aux1'
			gen typi=X`aux9'
			gen fretypi=X`aux10'
			quietly interval regu freregu
			quietly interval typi fretypi
			gen pD=regut*(regut>0)+typit*(typit>0)
			gen rD=D*inte*(inte>0)
			gen invrealest=1 if X`var'==78 & nnresre!=0
			replace invrealest=0 if X`var'==78 & nnresre==0
			replace invrealest=0 if X`var'!=78


			*case1: change in debt = D
			gen dcloan`var'=D if tB==`year' & D>0 & D!=. & invrealest==0
			*case2: change in debt = rD-pD
			replace dcloan`var'=rD-pD if tB<`year' & (B>D | B<D) & D>0 & D!=. & invrealest==0
			*case3: change in debt = 0
			replace dcloan`var'=0 if tB<`year' & B==D & D>0 & D!=. & invrealest==0
			*all which do not have such a loan
			replace dcloan`var'=0 if D==0
			replace dcloan`var'=0 if invrealest==1 & D>0 & D!=.

			noi count if dcloan`var'==.
			drop pD rD invrealest B D tB inte regu freregu typi fretypi regut typit

		}

		gen dcloan=dcloan2710+dcloan2727+dcloan2810+dcloan2827+dcloan2910+dcloan2927
	}
****************************************************compute changes for car (2208/2308/2408/7157) & vehicle loans (2509/2609)
	noi disp "		car loans"
	*years prior to 1992 were excluded because naming of variables in the codebooks 1989 and 1992 is slightly different, thus can't loop over
	*since change of primary mortgage not defined prior to 1995 won't use 1989 and 1992 anyway
	if `year'>=1995 {
		gen XX2212=X7537
		gen XX2312=X7536
```

```
gen XX2412=X7535
gen XX2513=X7531
gen XX2613=X7530

foreach var in 2208 2308 2408 2509 2609 {

        local aux1=`var'+1
        local aux4=`var'+4
        local aux5=`var'+5
        local aux6=`var'+6
        local aux7=`var'+7
        local aux10=`var'+10
        local aux11=`var'+11

        gen B=X`aux1'
        gen D=X`aux10'
        gen tB=X`var'
        gen inte=X`aux11'/10000
        gen regu=X`aux5'
        gen freregu=XX`aux4'
        gen typi=X`aux6'
        gen fretypi=X`aux7'
        quietly interval regu freregu
        quietly interval typi fretypi
        gen pD=regut*(regut>0)+typit*(typit>0)
        gen rD=D*inte*(inte>0)

        *case1: change in debt = D
        gen dcarloan`var'=D if tB==`year' & D>0 & D!=.
        *case2: change in debt = rD-pD
        replace dcarloan`var'=rD-pD if tB<`year' & (B>D | B<D) & D>0 & D!=.
        *case3: change in debt = 0
        replace dcarloan`var'=0 if tB<`year' & B==D & D>0 & D!=.
        *all which do not have a mortgage
        replace dcarloan`var'=0 if D==0

        noi count if dcarloan`var'==.
        drop pD rD B D tB inte regu freregu typi fretypi regut typit
```

```
		}
}
if `year'>=1995 {
	foreach var in  7157 {

		local aux1=`var'+1
		local aux5=`var'+5
		local aux6=`var'+6
		local aux7=`var'+7
		local aux8=`var'+8
		local aux12=`var'+12
		local aux13=`var'+13

		gen B=X`aux1'
		gen D=X`aux12'
		gen tB=X`var'
		gen inte=X`aux13'/10000
		gen regu=X`aux5'
		gen freregu=X`aux6'
		gen typi=X`aux7'
		gen fretypi=X`aux8'
		quietly interval regu freregu
		quietly interval typi fretypi
		gen pD=regut*(regut>0)+typit*(typit>0)
		gen rD=D*inte*(inte>0)

		*case1: change in debt = D
		gen dcarloan`var'=D if tB==`year' & D>0 & D!=.
		*case2: change in debt = rD-pD
		replace dcarloan`var'=rD-pD if tB<`year' & (B>D | B<D) & D>0 & D!=.
		*case3: change in debt = 0
		replace dcarloan`var'=0 if tB<`year' & B==D & D>0 & D!=.
		*all which do not have a mortgage
		replace dcarloan`var'=0 if D==0

		noi count if dcarloan`var'==.
		drop pD rD B D tB inte regu freregu typi fretypi regut typit
```

```
			}
			gen dcarloan=dcarloan2208+dcarloan2308+dcarloan2408+dcarloan7157+dcarloan2509+dcarloan2609
		}
******************************************education loans
		noi disp "		education loans"
		if `year'>1989 {
			foreach var in 7804 7827 7850 7904 7927 7950 {

				local aux1=`var'+1
				local aux2=`var'+2
				local aux11=`var'+11
				local aux12=`var'+12
				local aux13=`var'+13
				local aux14=`var'+14
				local aux18=`var'+18
				local aux20=`var'+20

				gen B=X`aux1'
				gen D=X`aux20'
				gen tB=X`var'
				gen inte=X`aux18'/10000
				gen regu=X`aux11'
				gen freregu=X`aux12'
				gen typi=X`aux13'
				gen fretypi=X`aux14'
				quietly interval regu freregu
				quietly interval typi fretypi
				gen pD=regut*(regut>0)+typit*(typit>0)
				gen rD=D*inte*(inte>0)
				gen pay=X`aux2'

				*case1: change in debt = D
				gen deduloan`var'=D if tB==`year' & D>0 & D!=.
				*case2: change in debt = rD-pD
				replace deduloan`var'=rD-pD if tB<`year' & (B>D | B<D) & D>0 & D!=.
				*case3: change in debt = 0
				replace deduloan`var'=0 if tB<`year' & B==D & D>0 & D!=.
				*all which do not have a mortgage
```

```
				replace deduloan`var'=0 if D==0

				noi count if deduloan`var'==.
				drop pD rD B D tB inte regu freregu typi fretypi regut typit pay

			}
		gen deduloan=deduloan7804+deduloan7827+deduloan7850+deduloan7904+deduloan7927+deduloan7950
		}
*****************************************************change in other debt
		noi disp "		other loans"
		if `year'>1989 {
			gen XX1038=X7567
			gen XX1209=X7565

			foreach var in 1034 1205 {

				local aux1=`var'+1
				local aux4=`var'+4
				local aux5=`var'+5
				local aux6=`var'+6
				local aux7=`var'+7
				local aux10=`var'+10
				local aux11=`var'+11

				gen B=X`aux1'
				gen D=X`aux10'
				gen tB=X`var'
				gen inte=X`aux11'/10000
				gen regu=X`aux5'
				gen freregu=XX`aux4'
				gen typi=X`aux6'
				gen fretypi=X`aux7'
				quietly interval regu freregu
				quietly interval typi fretypi
				gen pD=regut*(regut>0)+typit*(typit>0)
				gen rD=D*inte*(inte>0)

				*case1: change in debt = D
```

```
				gen dothloan`var'=D if tB==`year' & D>0 & D!=.
				*case2: change in debt = rD-pD
				replace dothloan`var'=rD-pD if tB<`year' & (B>D | B<D) & D>0 & D!=.
				*case3: change in debt = 0
				replace dothloan`var'=0 if tB<`year' & B==D & D>0 & D!=.
				*all which do not have a mortgage
				replace dothloan`var'=0 if D==0

				noi count if dothloan`var'==.
				drop pD rD B D tB inte regu freregu typi fretypi regut typit

			}
		gen dothloan=dothloan1034+dothloan1205
		}
***************************************keep relevant variables
		if `year'==1989 {
			keep Y1 YY1 homeeq othloc mrthel resdbt ccbal install odebt debt networth pirtotal ///
			X401 turndown X408 feardenial turnfear BNKRUPLAST5 PLOAN*
		}
		else if `year'==1992 {
			keep Y1 YY1 homeeq othloc mrthel resdbt ccbal install odebt debt networth pirtotal ///
			X401 turndown X408 X7586 feardenial turnfear X7583 BNKRUPLAST5 PLOAN*
		}
		else {
			keep Y1 YY1 dmor1 dmor domor dcloan ///
			dcarloan deduloan dothloan homeeq HEXTRACT_EVER othloc mrthel resdbt ccbal install odebt debt networth pirtotal ///
			X401 X7131 turndown X408 X7586 feardenial turnfear X7583 BNKRUPLAST5 PLOAN*
			gen ddebt2=dmor+domor+dcloan+dcarloan+deduloan+dothloan
		}
		gen year=`year'
		save "debt`year'.dta" , replace
}
*

*****combine all***********
clear
use "debt1989.dta"
append using "debt1992.dta"
```

```
append using "debt1995.dta"
append using "debt1998.dta"
append using "debt2001.dta"
append using "debt2004.dta"
append using "debt2007.dta"
append using "debt2010.dta"
append using "debt2013.dta"
*generate CPI
gen cpi=1 if year==2013
replace cpi=0.9331006399 if year==2010
replace cpi=0.8906340896 if year==2007
replace cpi=0.810936591 if year==2004
replace cpi=0.7614892379 if year==2001
replace cpi=0.6995346131 if year==1998
replace cpi=0.6588132635 if year==1995
replace cpi=0.6154741129 if year==1992
replace cpi=0.5532286213 if year==1989
*deflate everything
mdesc
replace ddebt2=ddebt2/cpi
replace dmor1=dmor1/cpi
replace dmor=dmor/cpi
replace domor=domor/cpi
replace dcloan=dcloan/cpi
replace dcarloan=dcarloan/cpi
replace deduloan=deduloan/cpi
replace dothloan=dothloan/cpi
*finish
gen imp=Y1-10*YY1
save "combined_allyears_debt.dta", replace
foreach year in 1989 1992 1995 1998 2001 2004 2007 2010 2013 {
	erase "combined`year'assets.dta"
}
*erst jetzt zum schluss alle Teildatensätze löschen
foreach year in 1989 1992 1995 1998 2001 2004 2007 2010 2013 {
	erase "debt`year'.dta"
}
*
```

```
*****************************************************************************************
****************Section 5: Combine everything into a general data set******************
*****************************************************************************************

*set directory
cd "C:\Users\Rafael\Google Drive\Data\SCF Data\original files"
*cd "H:\Data\SCF - Home Drive\original files" //in remote or Kingston desktop mode
*combine
clear
use "combined_allyears_debt.dta"
merge 1:1 Y1 year using "combined_allyears_assets.dta"
drop _merge
merge 1:1 Y1 year using "combined_allyears_demo.dta"
drop _merge
merge 1:1 Y1 year using "combined_allyears_income.dta"
drop _merge
sort year Y1
*cd "H:\Data\SCF - Home Drive" //in remote or Kingston desktop mode
*save "complete dataset v7.dta", replace
*save "complete dataset v8.dta", replace // v8 is with new definition of high quality indicator
*save "complete dataset v9.dta", replace // v9 is based on v8 and includes the variables for a sex/single household dummy
*save "complete dataset v10.dta", replace // v10 is based on newest definitions of d.D
cd "C:\Users\Rafael\Google Drive\Data\SCF Data\original files"
save "complete dataset v11.dta", replace // based on v10 plus updated HW and REP definitions


count if ddebt2>0 & ddebt2!=. //24,455 in complete dataset v11
count if ddebt2>0 & ddebt2!=. //22,786 in estimation data v7
count if REP>0 & REP!=. //7,493 in complete dataset v11
count if REP>0 & REP!=. //6,996 in estimaiton data v7


*****************************************************************************************
```

```
*********************Section 6: Set up imputation structure**************************
*************************************************************************************
*2min

****First: check mi structure
*mi describe
*mi varying

****Second: split data in seperate files for each implicate and "simulate" implicate 0
*code used from ÖNB Methodenband zu HFCS
clear
cd "C:\Users\Rafael\Google Drive\Data\SCF Data\original files"
*cd "H:\Data\SCF - Home Drive" //in remote or Kingston desktop mode
*create temporary implicate data sets
forvalues j=1(1)5 {
	clear
	use "complete dataset v11.dta"
	drop if year==1989 | year==1992
	keep if imp==`j'
	save "C:\Users\Rafael\Google Drive\Data\SCF Data\original files\temp`j'.dta", replace
}
*Create the zero implicate to simulate the original data
*Use one implicate of the data
use "C:\Users\Rafael\Google Drive\Data\SCF Data\original files\temp1.dta", clear
*Replace the implicate number by "0" to simulate the original data
replace imp=0
*Append all other implicates
forvalues i=1(1)5 {
	append using "C:\Users\Rafael\Google Drive\Data\SCF Data\original files\temp`i'.dta"
	erase "C:\Users\Rafael\Google Drive\Data\SCF Data\original files\temp`i'.dta"
}
*mdesc
tab year
*reconstruct M0 based on varying observations across the implicates
*laptop: takes more 10min
*desktop: 2min
global IMPUTED=""
local iter=1
```

```
foreach var of varlist X* turndown feardenial turnfear BNKRUPLAST5 mrthel homeeq HEXTRACT_EVER othloc resdbt ccbal install ///
odebt d* networth pirtotal PLOAN* liq cds nmmf stocks bond retqliq savbnd cashli othma othfin equitinc fin vehic ///
houses housecl oresre nnresre bus othnfin nfin asset REP_houses capcor_houses REP_oresre capcor_oresre REP HW hhsex age educ edcl ///
married kids lf race* OCCAT* indcat income ///
wageinc bussefarminc intdivinc kginc ssretinc transfothinc penacctwd norminc late LATE60 wsaved saved spendmor   ///
tpay mortpay conspay revpay pirmort pircons pirrev PIR40 ///
kghouse kgore kgbus FARMBUS_KG kgstmf kgtotal  {
		tempvar sd
		quietly bysort year YY1 : egen `sd'=sd(`var')
		quietly count if `sd'>0.00000001 & imp==0
		if r(N)>0 global IMPUTED "$IMPUTED `var'"
		quietly replace `var'=. if `sd'>0.00000001 & imp==0
		drop `sd'
		local iter=`iter'+1
		disp round((`iter'/172)*100), _continue
}
*
disp "$IMPUTED"

*************************
***NOTE: Stata mi manual p.85: there is no statistical distinction between passive and imputed variables!
*************************


****************************************************register everything as imputed*****************
****Third: Import as multiply imputed data
*Import the imputation structure of the data into Stata
mi import flong, m(imp) id(year YY1) clear
drop imp Y1
*Register the variables that are imputed
mi register imputed ///
X401 X408 X7586 X7583 X7131 X607 X614 X616 X617 X626 X627 X630 X631 X716 X717 X720 X1703 X1705 X1706 X1708 X1709 X1710 ///
X1803 X1805 X1806 X1808 X1809 X1810 X1903 X1905 X1906 X1908 X1909 X1910 X3110 X3210 X3310 X5804 X5805 X5809 X5810 ///
X5814 X5815 X301 X4511 X5111 X6536 X6770 X6780 X6781 X6784 X6785 X7017 X304 X407 X3004 X3005 X7509 X7507 X7650 ///
turndown feardenial turnfear BNKRUPLAST5 mrthel homeeq HEXTRACT_EVER othloc resdbt ccbal install odebt debt dmor1 dmor ///
domor dcloan dcarloan deduloan dothloan ddebt2 networth pirtotal PLOAN* liq cds nmmf stocks bond retqliq savbnd cashli othma ///
othfin equitinc fin vehic houses oresre nnresre bus othnfin nfin asset age educ edcl lf race* OCCAT* indcat income ///
REP_houses capcor_houses REP_oresre capcor_oresre REP HW ///
```

```
wageinc bussefarminc intdivinc kginc ssretinc transfothinc penacctwd norminc wsaved saved spendmor late LATE60 ///
tpay mortpay conspay revpay pirmort pircons pirrev PIR40 ///
kghouse kgore kgbus FARMBUS_KG kgstmf kgtotal
*so out of 34,479obs per implicate 33,842 are incomplete and thus imputed (i.e. 98%)

*correct potential problem
mi unregister PLOAN8 X7017
mi register regular PLOAN8 X7017
mi varying
*NOTE: the fact that Y1 and imp are super varying is clear. By definition there are no missing
*		values for these variables in m=0 but since they either denote the implicate or have the
*		implicate implicite in, they are varying across implicates. No need to worry about that.
*		wgt is supervarying because there are no missing values in m=0 but weights are still different
*		in the other implicates. So that's just how it is, as long as wgt is not registered should be fine (i.e.
*		no automatic changes/"fixes" are applied).
*save "complete dataset v7 mi set.dta", replace
*save "complete dataset v8 mi set.dta", replace // version 8 includes new high quality definitions
*save "complete dataset v9 mi set.dta", replace // version 9 includes sex/single dummy information
*save "complete dataset v10 mi set.dta", replace // version 10 includes sex/single dummy information and won't include logs only ihs
*save "complete dataset v11 mi set.dta", replace // version 11 is based on updated d.D definitions
save "complete dataset v12 mi set.dta", replace // version 12 includeds updated REP and HW definitions



*****************************************************************************************
*******************Section 7: Creation of additional variables************************
*****************************************************************************************
*entire section 7 takes 13min

clear
cd "C:\Users\Rafael\Google Drive\Data\SCF Data\original files"
*cd "H:\Data\SCF - Home Drive" //in remote or Kingston desktop mode
use "complete dataset v12 mi set.dta", clear


*generate income measure without pension account withdrawals
*NOTE: In the SCF from 2004 onwards income includes a measure of withdrawals from pension
```

```
*               accounts named "penacctwd" in the summary data set. For consistency create an income
*               series for the years 2004-2013 where this component is not included!
mi passive: gen income2=income
mi passive: replace income2=income-penacctwd if year==2004 | year==2007 | year==2010 | year==2013

*drop if income2=0
*(one could also change to mi wide style but prefer this way, see first round code)
count if income2==0 & _mi_m!=0 // 1409
gen dum=1 if income2==0
bysort year YY1: egen dum1=total(dum)
*drop observation if income or assets are zero in any implicate
drop if dum1>0 //1692
drop dum dum1

*now define a new measure of change in debt which excludes lines of
*credit (deltaloc, based on X1108 for example) and also excludes credit
*cards (dccbal, based on X413 for example).
*The reason is that there is not adequate timing information availabe for
*these two items.
*The old measure: ddebt=dmor+deltaloc+domor+dcloan+dccbal+dcarloan+deduloan+dothloan
*The new measure: ddebt2=dmor+domor+dcloan+dcarloan+deduloan+dothloan
*construct new D(-1)
        mi passive: gen debtlag2=debt-ddebt2
        *now there are some (2793) observations with debtlag2<0
        count if debtlag2<0 & debtlag2!=. //2793
        count if debtlag2<0 & _mi_m!=0 & debtlag2!=. //2497
        count if debtlag2<0 //2793
        *drop these observations
        gen dum=1 if debtlag2<0
        bysort year YY1: egen dum1=total(dum)
        sum dum1,d
        count if dum1>0 //3600
        drop if dum1>0
        drop dum dum1

*generate some necessary additional (i.e. passive) variables
mi passive: gen deltad2y=(ddebt2/income2)
mi passive: gen age2=age*age
```

```
mi passive: gen debty=(debt/income2)

*create dummy for less than normal income
	mi passive: gen lowinc=1 if X7650==2
	mi passive: replace lowinc=0 if X7650==1 | X7650==3
*create additional housing variables
	*dummy for possessing residential real estate
		mi passive: gen dumHW=1 if HW>0 & HW!=.
		mi passive: replace dumHW=0 if HW==0
	*dummy for purchasing in the current period
		mi passive: gen dREP=1 if REP>0 & REP!=.
		mi passive: replace dREP=0 if REP==0
	*define capital gains measure that only applies to primary residence and oresre
		*see v91 for more detailed checks
			mi passive: gen kgoresre=kgore if oresre>0 & oresre!=.
			mi passive: replace kgoresre=0 if oresre==0 & kgore!=.
		*now combine both measures
			mi passive: gen kgHW=kghouse+kgoresre
		*define dummy which equals 1 if there is some house or oresre (not just if kgHW=0)
			mi passive: gen dkgHW=1 if kgHW!=. & ((oresre>0 & oresre!=.) | (houses>0 & houses!=.))
			mi passive: replace dkgHW=0 if kgHW==0 & oresre==0 & houses<=0
*generate dummies for debt and financial wealth
	mi passive: gen dfin=1 if fin>0 & fin!=.
	mi passive: replace dfin=0 if fin==0
	mi passive: gen ddebt2lag=1 if debtlag2>0 & debtlag2!=.
	mi passive: replace ddebt2lag=0 if debtlag2==0
*year dummies
forvalues year=1995(3)2013 {
	gen dum`year'=1 if year==`year'
	replace dum`year'=0 if dum`year'==.
}
*generate a kids dummy (not varying so no mi command necessary)
	gen kidsd=1 if kids>0 & kids<=10
	replace kidsd=0 if kids==0
*generate broad and strong credit constraint dummies
	mi passive: gen credcons=1 if X407==1 | X407==3
	mi passive: replace credcons=0 if X407==5 | X407==0
	mi passive: gen credconsstr=1 if X407==1
```

```
	mi passive: replace credconsstr=0 if X407==5 | X407==0 | X407==3
*check structure before continuing
	mi register regular kidsd
	mi register imputed X30022 X30074
	mi varying
*save "final data set estimation v1.dta", replace
*save "final data set estimation v2.dta", replace //includes new high quality dummies
*save "final data set estimation v3.dta", replace //based on v2 but includes new ddebt definition without credit linse and credit card balances
*save "final data set estimation v4.dta", replace //try to create something like v6 where the quantile regression worked
*save "final data set estimation v9.dta", replace //based on v3 and includes more household dummies (sex, single household)
*save "final data set estimation v10.dta", replace //based on v3 and no logs any more
*save "complete dataset v10 mi set.dta", replace // version 10 includes sex/single dummy information and won't include logs only ihs
*save "complete dataset v11 mi set.dta", replace // newest definitions of d.D
save "complete dataset v12 mi set.dta", replace // updated versions of HW and REP and fewer unused variables




***************************************************************************************************
*********************Section 8: Define Relative Income Measures***********************
***************************************************************************************************

*use data created in section 7
	clear all
	cd "C:\Users\Rafael\Google Drive\Data\SCF Data\original files"
	*cd "H:\Data\SCF - Home Drive" //in remote or Kingston desktop mode
	use "complete dataset v12 mi set.dta", replace
*fix the weights
	total wgt if year==2010 & _mi_m==3
	replace wgt=wgt*5
*create income percentiles
	gen incperc=.
	forvalues year=1995(3)2013 {
		disp `year'
		forvalues imp=1(1)5 {
			xtile incperc`year'imp`imp'=income2 if year==`year' & _mi_m==`imp' [pweight=wgt], n(100)
			replace incperc=incperc`year'imp`imp' if year==`year' & _mi_m==`imp'
```

```
                drop incperc`year'imp`imp'
        }

    }
    sum incperc if year==2004 & _mi_m==3 ,d
    tab incperc if year==2004 & _mi_m==3
*create robustness check for head
    gen head_incperc=1-incperc/100

**********top income percentile as cut-off (incpXre) and average top income (avinctopXre)***************
    gen wgtround=round(wgt)
    mi passive: gen eduaux=1 if edcl==4
    mi passive: replace eduaux=0 if edcl<4
    sum eduaux edcl
    *race // 1=white-non-Hispanic 2=black 3=hispanic 5=other
    mi passive: gen raceaux=1 if race==1
    mi passive: replace raceaux=2 if race==2
    mi passive: replace raceaux=3 if race==3
    mi passive: gen raceaux2=1 if race==1
    mi passive: replace raceaux2=2 if race==2
    count if raceaux==.
    count if race==.
*min income of first obs in top 5/10/20% as reference for all others
*and also average income of top 20 10 5% as reference for all others
*again use education and race groupings
    gen incp99re=.
    gen incp95re=.
    gen incp90re=.
    gen incp80re=.
    gen avinctop1re=.
    gen avinctop5re=.
    gen avinctop10re=.
    gen avinctop20re=.
    gen top1re=.
    gen top5re=.
    gen top10re=.
    gen top20re=.
    foreach quant in 1 5 10 20 {
```

```
		mat perc`quant'=J(30,7,1)
		mat av`quant'=J(30,7,1)
		mat share`quant'=J(30,7,1)
		mat obstot`quant'=J(30,7,1)
		mat obstop`quant'=J(30,7,1)
		mat weight`quant'=J(30,7,1)
}
local iter=0
qui forvalues year=1995(3)2013 {
		nois disp "					`year'"
		local iter=`iter'+1
		local iter2=0
		forvalues i=1/5 {
			nois disp "		`i'"
			capture drop group
			egen group=group(eduaux raceaux) if year==`year' & _mi_m==`i'
			forvalues j=1(1)6 {
				local iter2=`iter2'+1
				foreach quant in 1 5 10 20 {
					local num=100-`quant'
					*the percentile measure
						_pctile income2 [pweight=wgt] if group==`j', p(`num')
						local cut=r(r1)
						mat perc`quant'[`iter2',`iter']=`cut'
						*replace incp`num're=r(r1) if group==`j' & income2<`cut'
					*the average measure
						*sum income2 [fw=wgtround] if group==`j' & income2>=`cut'
						*mat av`quant'[`iter2',`iter']=r(mean)
						*replace avinctop`quant're=r(mean) if group==`j' & income2<`cut'
					*the top income share measure
						total income2 [fw=wgtround] if group==`j'
						mat tot=e(b)
						total income2 [fw=wgtround] if group==`j' & income2>=`cut'
						mat share`quant'[`iter2',`iter']=e(b)/tot[1,1]
						local share=share`quant'[`iter2',`iter']
						replace top`quant're=`share' if group==`j' & income2<`cut'
					*observations and mean weight for the groups
						*sum wgt if group==`j' & income2>=`cut'
```

```
								*mat weight`quant'[`iter2',`iter']=r(mean)
								*mat obstop`quant'[`iter2',`iter']=r(N)
								*sum income if group==`j'
								*mat obstot`quant'[`iter2',`iter']=r(N)
					}
				}
		}
}
foreach var in perc av share obstot obstop weight {
		foreach quant in 1 5 10 20 {
				mat colnames `var'`quant'=1995 1998 2001 2004 2007 2010 2013
				mat rownames `var'`quant'=i1-nocol|w i1-nocol|b i1-nocol|h i1-col|w i1-col|b i1-col|h ///
															i2-nocol|w i2-nocol|b i2-nocol|h i2-col|w i2-col|b i2-col|h ///
															i3-nocol|w i3-nocol|b i3-nocol|h i3-col|w i3-col|b i3-col|h ///
															i4-nocol|w i4-nocol|b i4-nocol|h i4-col|w i4-col|b i4-col|h ///
															i5-nocol|w i5-nocol|b i5-nocol|h i5-col|w i5-col|b i5-col|h
		}
}
*reorganize the matrices
foreach var in perc av share obstot obstop weight {
		foreach quant in 1 5 10 20 {
				forvalues i=1/6 {
						mat `var'`quant'aux`i'=(`var'`quant'[`i',1..7] \ `var'`quant'[`i'+6,1..7] \ `var'`quant'[`i'+12,1..7] \ `var'`quant'[`i'+18,1..7] \ `var'`quant'[`i'+24,1..7])
						if `i'==1 mat `var'`quant'new=`var'`quant'aux1
						if `i'>1 mat `var'`quant'new=(`var'`quant'new \ `var'`quant'aux`i')
				}
				mat list `var'`quant'new
		}
}
*export the matrices
cd "C:\Users\Rafael\Google Drive\Data\SCF Data\descriptive plots\income\inequality by groups"
foreach var in perc av share obstot obstop weight {
		foreach quant in 1 5 10 20 {
				xml_tab `var'`quant'new, save("`var'`quant'.xml") replace
		}
}
*run some checks
		count if incp96re==.
```

```
                    count if avinctop5re==.
                    sum incp96re avinctop5re
                    order income2, last
                    tab group
                    count if income2>avinctop5re
                    *check percentiles and average incomes
                              _pctile income2 [fw=wgtround] if year==2004 & _mi_m==3 & eduaux==1 & race==3, p(99)
                              disp `r(r1)'
                              mat list perc1new
                              sum income2 [fw=wgtround] if year==2004 & _mi_m==3 & eduaux==1 & race==3 & income2>=r(r1)
                              mat list av1new
                    *check out shares
                              _pctile income2 [fw=wgtround] if year==2004 & _mi_m==3 & eduaux==1 & race==3, p(99)
                              disp `r(r1)'
                              total income2 [fw=wgtround] if year==2004 & _mi_m==3 & eduaux==1 & race==3 & income2>=`r(r1)'
                              mat test=e(b)
                              total income2 [fw=wgtround] if year==2004 & _mi_m==3 & eduaux==1 & race==3
                              mat test2=e(b)
                              mat test3=test[1,1]/test2[1,1]
                              mat list test3
                              mat list share1new
*Now compute them based on the regional identifier X30074 for 1995 and 1998 when it is available.
*This will allow me to run robustness checks with respect to what is the impact of not having
*regional identifiers. I can then also run robustness check based on including region dummies!
          gen incp99regio=.
          gen incp95regio=.
          gen incp90regio=.
          gen incp80regio=.
          gen avinctop1regio=.
          gen avinctop5regio=.
          gen avinctop10regio=.
          gen avinctop20regio=.
          qui forvalues year=1995(3)1998 {
                    nois disp "                              `year'"
                    forvalues i=1/5 {
                              nois disp "          `i'"
                              foreach quant in 1 5 10 20 {
                                        local num=100-`quant'
```

```
					forvalues j=1/9 {
						*the percentile measure
						_pctile income2 [pweight=wgt] if year==`year' & _mi_m==`i' & X30074==`j', p(`num')
						local cut=r(r1)
						replace incp`num'regio=r(r1) if year==`year' & _mi_m==`i' & X30074==`j' & income2<`cut'
						*the average measure
						sum income2 [fw=wgtround] if year==`year' & _mi_m==`i' & X30074==`j' & income2>=`cut'
						replace avinctop`quant'regio=r(mean) if year==`year' & _mi_m==`i' & X30074==`j' & income2<`cut'
					}
				}
			}
		}
		*basic checks
		sum avinctop5regio X30074
*next highest decile with finer split of top decile
		gen refincA=.
		qui forvalues year=1995(3)2013 {
			nois disp "						`year'"
			local im 5 //set the number of implicates for which the computation is done
			forvalues i=1/`im' {

				noi disp "implicate `i'"
				noi disp "top1"
				qui sum income2 [fw=wgtround] if year==`year' & _mi_m==`i' & incperc==100
				replace refincA=r(mean) if year==`year' & _mi_m==`i' & incperc<=99 & incperc>=96

				noi disp "99-96"
				qui sum income2 [fw=wgtround] if year==`year' & _mi_m==`i' & incperc<=99 & incperc>=96
				replace refincA=r(mean) if year==`year' & _mi_m==`i' & incperc<=95 & incperc>=91

				noi disp "95-91"
				qui sum income2 [fw=wgtround] if year==`year' & _mi_m==`i' & incperc<=95 & incperc>=91
				replace refincA=r(mean) if year==`year' & _mi_m==`i' & incperc<=90 & incperc>=81

				noi disp "deciles"
				forvalues quant=90(-10)20 { //Deciles 9 to 1
					qui sum income2 [fw=wgtround] if year==`year' & _mi_m==`i' & incperc<=`quant' & incperc>=`quant'-9
					replace refincA=r(mean) if year==`year' & _mi_m==`i' & incperc<=`quant'-10 & incperc>=`quant'-19
```

```
				}
			}
		}
		*run some checks
		sum income2 [fw=wgtround] if year==2004 & _mi_m==1 & incperc<=70 & incperc>=61
*check out distribution of refincA
		gen test=refincA/income2
		foreach year in 1995 {
			sum test if incperc<=99 & incperc>=96 & year==`year' & _mi_m==3
			sum test if incperc<=95 & incperc>=91 & year==`year' & _mi_m==3
			sum test if incperc<=90 & incperc>=81 & year==`year' & _mi_m==3
			sum test if incperc<=80 & incperc>=71 & year==`year' & _mi_m==3
			sum test if incperc<=70 & incperc>=61 & year==`year' & _mi_m==3
			sum test if incperc<=20 & incperc>=11 & year==`year' & _mi_m==3
			sum test if incperc<=10 & incperc>=1 & year==`year' & _mi_m==3
		}
		hist test if incperc<=99 & incperc>=96 & year==2004 & _mi_m==3
		hist test if incperc<=95 & incperc>=91 & year==2004 & _mi_m==3
		hist test if incperc<=90 & incperc>=81 & year==2004 & _mi_m==3
		hist test if incperc<=80 & incperc>=71 & year==2004 & _mi_m==3
		hist test if incperc<=70 & incperc>=61 & year==2004 & _mi_m==3
		hist test if incperc<=20 & incperc>=11 & year==2004 & _mi_m==3
		hist test if incperc<=10 & incperc>=1 & year==2004 & _mi_m==3
		sum test if incperc<=10 & incperc>=4 & year==1995 & _mi_m==3
		drop test


**********headcount no grouping (head) and with edu-race grouping (headre)***************
*define head count measure (three versions)
		*version 1: only headcount of richer individuals (which is equivalent to a finer definition of incperc/100) in each year
		*version 2: define headcount but for regional identifiers (only possible for 1995 and 1998 when such identifiers are available
		*version 3: headcount with education (college degree vs no degree) and race (black, white, hispanic, other) grouping in each year
*version 1
		*make sure grouping variables works
			qui forvalues year=1995(3)1995 {
				forvalues i=1/1 {
					capture drop rankinc
					capture drop group
```

```
				capture drop group2
				egen rankinc = rank(income2) if year==`year' & _mi_m==`i', t //lowest value is ranked 1
				egen group = group(rankinc) if year==`year' & _mi_m==`i'
				egen group2 = group(income2) if year==`year' & _mi_m==`i'
				gen test=group/group2
				noi sum test, mean
				noi disp "`r(mean)'"
				drop test
			}
		}
*compute number richer households
*35 year-implicate "cells", takes almost 2.5h on laptop (49 min desktop)
		gen head=.
		qui forvalues year=1995(3)2013 {
		noi disp " `year'"
			forvalues i=1/5 {
				noi disp "`i'"
				capture drop group
				egen group = group(income2) if year==`year' & _mi_m==`i' //lowest value/group is number 1
				sum group
				local max=r(max) //gives the number of distinct observation for that year-implicate
				sum income2 [fw=wgtround] if group!=.
				local tot=r(N) //gives the number of people for that year-implicate
				forvalues j=1(1)`max' {
					sum income2 [fw=wgtround] if group>`j' & group!=.
					replace head=(r(N))/`tot' if year==`year' & _mi_m==`i' & group==`j'
				}
			}
		}
*run some checks
		sum head group
		count if year==2013 & _mi_m==5
		sum income2 if group!=.
		*count if
		gen test=(1-head)*100/incperc
		sum test,d
		order head head_incperc incperc test, last
		sum incperc if test<0.50 //all the very small values are in the first percentile
```

```
			sum incperc if test<0.80 //up to 4th percentile
			sum incperc if test<0.97 //up to 33rd percentile
			sum incperc if test>1.2 & test!=. //up to 7th percentile
			sum incperc if test>1.15 & test!=. //up to 11th percentile
			sum incperc if test>1.1 & test!=. //up to 27th percentile
			sum incperc if test>1.05 & test!=. //up to 38th percentile
			*The difference for lower income household most likely comes from different treatments of ties in the pctile command
			*and the way head is computed. Lower income households have higher weights thus different treatment then materializes
			*in a stronger way.
			*compare with head measure based on incperc
			gen test=head/head_incperc
			sum test,d
*version 2		(26 min desktop)
		gen headregio=.
			qui forvalues year=1995(3)1998 {
			noi disp " `year'"
				forvalues i=1/5 {
					noi disp "`i'"
					forvalues j=1(1)9 {
						capture drop group
						egen group = group(income2) if year==`year' & _mi_m==`i' & X30074==`j' //lowest value/group is number 1
						sum group
						local max=r(max) //gives the number of distinct observation for that year-implicate
						sum income2 [fw=wgtround] if group!=.
						local tot=r(N) //gives the number of people for that year-implicate
						forvalues j=1(1)`max' {
							sum income2 [fw=wgtround] if group>`j' & group!=.
							replace headregio=(r(N))/`tot' if year==`year' & _mi_m==`i' & group==`j'
						}
					}
				}
			}
		*run some checks
		sum headregio
*version 3: also include race and edu classifications
*edcl==4 //college degree
		gen headre=.
		qui forvalues year=1995(3)2013 {
```

```
noi disp "              `year'"
        forvalues i=1/5 {
                noi disp " imp `i'"
                capture drop group
                egen group=group(eduaux raceaux) if year==`year' & _mi_m==`i'
                forvalues j=1(1)6 {
                        capture drop group2
                        egen group2 = group(income2) if group==`j' //lowest value/group is number 1
                        sum group2
                        local max=r(max) //number of distinct income values
                        noi disp "`max'"
                        sum income2 [fw=wgtround] if group2!=.
                        local tot=r(N) //number of households in subsample
                        noi disp "`tot'"
                        forvalues h=1(1)`max' {
                                sum income2 [fw=wgtround] if group2>`h' & group2!=.
                                replace headre=(`r(N)')/`tot' if group2==`h'
                        }
                }
        }
}
*run some checks
        *group 1: no college and white
        *group 2: no college and black
        *group 3: no college and hispanic
        *group 4: college and white
        *group 5: college and black
        *group 6: college and hispanic
_pctile income2 [pweight=wgt] if group==4, p(76)
disp "`r(r1)'" // 151k
capture drop group2
egen group2 = group(income2) if group==4
sum group2
sum income2 [fw=wgtround] if group2!=.
local tot=r(N)
sum income2 [fw=wgtround] if group2>382 & group2!=.
local aux=(`r(N)')/`tot'
disp "`aux'"
```

```
*check the data structure now
	mi varying
*create dummy for not being in bottom 3%
	gen dperc31=1 if incperc>3
	bysort year YY1: egen dperc3=total(dperc31)
	count if dperc3>4 //196,278
*for now save the data created up to this point
	*save "final data set estimation v5.dta", replace
	*save "final data set estimation v6.dta", replace //contains new definitions of high quality subsample and Y~ and C~
	*save "final data set estimation v7.dta", replace //based on v6 but with new change in debt measure excluding lines of credit and cred cards
	*save "final data set estimation v8.dta", replace //attempt to recreate v6
	*save "final data set estimation v9.dta", replace //new high quality debt measure and new demographic information and not taking logs of income scaled variables (peerinc and norminc(
	*save "final data set estimation v10.dta", replace //new high quality debt measure and new demographic information and only ihs transformation
	*save "final data set estimation v11.dta", replace //newest definition of d.D
	*save "final data set estimation v12.dta", replace //newest definitions of d.D, Y~ and d.HW
	*save "final data set estimation v13.dta", replace //newest definitions of d.D, Y~ and HW dropped vars not used
	cd "C:\Users\Rafael\Google Drive\Data\SCF Data"
	save "final data set estimation v14.dta", replace //based on v13 but headre is based on white, black and hispanic and with share measures



*****************************************************************************************
****************************Section 9: IHS transform the data*********************************
*****************************************************************************************

*use data created in section 8
	clear all
	cd "C:\Users\Rafael\Google Drive\Data\SCF Data"
	*cd "H:\Data\SCF - Home Drive" //in remote or Kingston desktop mode
	use "final data set estimation v13.dta", replace

*****************************IHS transform ********************

	foreach var income2 HW houses REP debtlag2 fin ///
	in incp99re incp95re incp90re incp80re ///
	incp99regio incp95regio incp90regio incp80regio ///
	avinctop1re avinctop5re avinctop10re avinctop20re ///
```

```
	avinctop1regio avinctop5regio avinctop10regio avinctop20regio ///
	refincA {
		disp "                    `var'"
		gen ihs_`var'=ln(`var'+sqrt(`var'^2+1))
	}
	mi register passive ihs_income2 ihs_fin ihs_debtlag2 ihs_houses ihs_REP ihs_HW
*check the data structure now
	mi varying
*for now save the data created up to this point
	*save "estimation data v1.dta", replace
	*save "estimation data v3.dta", replace //based on v1 but demeaned for 1998
	*save "estimation data v4.dta", replace //based on v3 but including demeaned debtlag
	*save "estimation data v5.dta", replace //newest d.D definitions
	*save "estimation data v6a.dta", replace //newest definitions of d.D, Y~ and d.HW
	*save "estimation data v6b.dta", replace //same as 6a but fixed the problems with avinctopXre
	*save "estimation data v7.dta", replace //same as 6b but includes HW which includes other residential real estate and new topY~ measures
	*save "estimation data v8.dta", replace //same as 7 but with changed percentile cut offs for Y~
	save "estimation data v9.dta", replace //same as 8 but ehadre is based on white black and hispanic and with share measures
```

Appendix VI

This Appendix contains the code used to produce the descriptive analysis for chapter 3.

```
*This file only creates the dataset based on which plots can be made. There is a seperate
*file (descriptive statistics quick plots) for the plots in order to keep the do files manageable.
	cd "C:\Users\Rafael\Google Drive\Data\SCF Data\original files"

*load and adopt the data
	clear all
	use "complete dataset v11.dta"
*drop the replicate weights
	drop repwgt*
*NOTE: first check whether the weights are correctly specified (about 100mio households
*	per year per implicate!
	total wgt if year==2010 & imp==3
	replace wgt=wgt*5

*set graph directory
	global graphout "C:\Users\Rafael\Dropbox\Diss\paper 4\graphs"



*generate income measure without pension account withdrawals
*NOTE: In the SCF from 2004 onwards income includes a measure of withdrawals from pension
*		accounts named "penacctwd" in the summary data set. For consistency create an income
*		series for the years 2004-2013 where this component is not included!
	gen income2=income
	replace income2=income-penacctwd if year==2004 | year==2007 | year==2010 | year==2013
*Now check the difference between income and income2
	gen inctest=income2/income if year==2004 | year==2007 | year==2010 | year==2013
	hist inctest if (year==2004 | year==2007 | year==2010 | year==2013) & inctest<0.9
	count if inctest<0.9
	sum inctest,d


*************************************************************************
*********************Picture 1: Debt to Income***********************
```

```
**********************************************************************

***Aggregate Total Income********************************************
	*compute total income per percentile per implicate and average across implicates
		gen totinc=income*wgt
	*compute total income first
		forvalues year=1989(3)2013 {
			local im 5 //number of implicates needs to be specified
			quie forvalues imp=1/`im' {
				total totinc if year==`year' & imp==`imp'
				mat tot`year'inc`imp' =(e(b))
				if `imp'==1 {
					local tot`year' "tot`year'inc`imp'"
				}
				else {
					local tot`year' "`tot`year''+tot`year'inc`imp'"
				}
			}
			disp "`tot`year''"
			qui mat tot`year'=(`tot`year'')/`im'
			qui mat colnames tot`year' = totinc
			qui mat rownames tot`year' = `year'
		}
	*create a year variable for plotting
		gen yearplot=1989 in 1
		replace yearplot=1992 in 2
		replace yearplot=1995 in 3
		replace yearplot=1998 in 4
		replace yearplot=2001 in 5
		replace yearplot=2004 in 6
		replace yearplot=2007 in 7
		replace yearplot=2010 in 8
		replace yearplot=2013 in 9
		label var yearplot " "
	*create an overal matrix with total income for all years and also a corresponding variable
		mat tot=(tot1989 \ tot1992 \ tot1995 \ tot1998 \ tot2001 \ tot2004 \ tot2007 \ tot2010 \ tot2013)
		mat list tot
		gen totalinc=.
```

```
			forvalues i=1/9 {
				replace totalinc=tot[`i',1]/10^9 in `i'
			}
		*plot total income
			twoway line totalinc yearplot, scheme(sj) ///
			legend(subtitle("")) ytit("bn. 2013 US$", height(8)) xlabel(1989 1992 1995 1998 2001 2004 2007 2010 2013) ///
			graphregion(fcolor(white)) title("aggregate income")
			graph export "$graphout\0 income level.tif", as(tif) replace

***Aggregate Total Debt*******************************************
		gen totdebt=debt*wgt
		forvalues year=1989(3)2013 {
			local im 5 //number of implicates needs to be specified
			quie forvalues imp=1/`im' {
				total totdebt if year==`year' & imp==`imp'
				mat tot`year'debt`imp' =(e(b))
				if `imp'==1 {
					local tot`year' "tot`year'debt`imp'"
				}
				else {
					local tot`year' "`tot`year''+tot`year'debt`imp'"
				}
			}
			disp "`tot`year''"
			qui mat tot`year'=(`tot`year'')/`im'
			qui mat colnames tot`year' = totdebt
			qui mat rownames tot`year' = `year'
		}
		*
		mat tot=(tot1989 \ tot1992 \ tot1995 \ tot1998 \ tot2001 \ tot2004 \ tot2007 \ tot2010 \ tot2013)
		mat list tot
		gen totaldebt=.
		forvalues i=1/9 {
			replace totaldebt=tot[`i',1]/10^9 in `i'
		}
		*plot aggregate debt
			twoway line totaldebt yearplot, scheme(sj) ///
			legend(subtitle("")) ytit("bn. 2013 US$", height(8)) xlabel(1989 1992 1995 1998 2001 2004 2007 2010 2013) ///
```

```
                graphregion(fcolor(white)) title("aggregate household liabilities")
                graph export "$graphout\0 debt level.tif", as(tif) replace
***Debt over income
        gen aggdebtinc=totaldebt/totalinc*100
        twoway line aggdebtinc yearplot, scheme(sj) ///
        legend(subtitle("")) ytit("% of aggregate income", height(8)) xlabel(1989 1992 1995 1998 2001 2004 2007 2010 2013) ///
        graphregion(fcolor(white)) title("aggregate debt to income ratio")
        graph export "$graphout\1 debt income ratio.tif", as(tif) replace


***Average Debt-to-Income*******************************************
        gen wgtround=round(wgt)
        gen dy=debt/income
        sum dy [fw=wgtround],d
        count if dy>100 & dy!=.
        sum dy [fw=wgtround] if dy<=100 & dy!=.,d
        replace dy=. if dy>100
        sum dy [fw=wgtround] if year==2004 & imp==3,d
        sum dy [fw=wgtround] if year==2004 & imp==3 & dy<50,d
        sum dy [fw=wgtround] if year==2004 & imp==3 & dy<25,d
        sum dy [fw=wgtround] if year==2004 & imp==3 & debt>0,d
        forvalues year=1989(3)2013 {
                local im 5 //number of implicates needs to be specified
                quie forvalues imp=1/`im' {
                        sum dy [fw=wgtround] if year==`year' & imp==`imp',d
                        mat dy`year'`imp' =r(p50)
                        if `imp'==1 {
                                local dy`year' "dy`year'`imp'"
                        }
                        else {
                                local dy`year' "`dy`year''+dy`year'`imp'"
                        }
                }
                disp "`dy`year''"
                qui mat dy`year'=(`dy`year'')/`im'
                qui mat colnames dy`year' = dy
                qui mat rownames dy`year' = `year'
        }
```

```
*
mat dy=(dy1989 \ dy1992 \ dy1995 \ dy1998 \ dy2001 \ dy2004 \ dy2007 \ dy2010 \ dy2013)
mat list dy
capture drop dyagg
gen dyagg=.
forvalues i=1/9 {
        replace dyagg=dy[`i',1]*100 in `i'
}
*plot aggregate debt
        twoway line dyagg yearplot, scheme(sj) ///
        legend(subtitle("")) ytit("% of household income", height(8)) xlabel(1989 1992 1995 1998 2001 2004 2007 2010 2013) ///
        graphregion(fcolor(white)) title("median debt to income ratio")
        graph export "$graphout\1.2 median debt to income.tif", as(tif) replace


***Average Debt-to-Income for debt holders only********************************************
sum dy,d
sum dy if debt>0 & debt!=.,d
forvalues year=1989(3)2013 {
        local im 5 //number of implicates needs to be specified
        quie forvalues imp=1/`im' {
                sum dy  [fw=wgtround] if year==`year' & imp==`imp' & debt>0 & debt!=.,d
                mat dypos`year'`imp' =r(p50)
                if `imp'==1 {
                        local dypos`year' "dypos`year'`imp'"
                }
                else {
                        local dypos`year' "`dypos`year''+dypos`year'`imp'"
                }
        }
        disp "`dypos`year''"
        qui mat dypos`year'=(`dypos`year'')/`im'
        qui mat colnames dypos`year' = dypos
        qui mat rownames dypos`year' = `year'
}
*
mat dypos=(dypos1989 \ dypos1992 \ dypos1995 \ dypos1998 \ dypos2001 \ dypos2004 \ dypos2007 \ dypos2010 \ dypos2013)
mat list dypos
```

```
	capture drop dyposagg
	gen dyposagg=.
	forvalues i=1/9 {
		replace dyposagg=dypos[`i',1]*100 in `i'
	}
	*plot aggregate debt
		twoway line dyposagg yearplot, scheme(sj) ///
		legend(subtitle("")) ytit("% of household income", height(8)) xlabel(1989 1992 1995 1998 2001 2004 2007 2010 2013) ///
		graphregion(fcolor(white)) title("median debt to income ratio, D>0 only")
		graph export "$graphout\1.3 median debt to income, debt holds only.tif", as(tif) replace


***Proportion of debt holders********************************************
	forvalues year=1989(3)2013 {
		local im 5 //number of implicates needs to be specified
		qui forvalues imp=1/`im' {
			total wgt if year==`year' & imp==`imp'
			mat totpop=e(b)
			total wgt if year==`year' & imp==`imp' & debt>0 & debt!=.
			mat debtors`year'`imp'=e(b)/totpop[1,1]
			if `imp'==1 {
				local debtors`year' "debtors`year'`imp'"
			}
			else {
				local debtors`year' "`debtors`year''+debtors`year'`imp'"
			}
		}
		disp "`debtors`year''"
		qui mat debtors`year'=(`debtors`year'')/`im'
		qui mat colnames debtors`year' = debtors
		qui mat rownames debtors`year' = `year'
	}
	*
	mat debtors=(debtors1989 \ debtors1992 \ debtors1995 \ debtors1998 \ debtors2001 \ debtors2004 \ debtors2007 \ debtors2010 \ debtors2013)
	mat list debtors
	gen debtors=.
	forvalues i=1/9 {
		replace debtors=debtors[`i',1]*100 in `i'
```

```
	}
	*plot aggregate debt
		twoway line debtors yearplot, scheme(sj) ///
		legend(subtitle("")) ytit("% of households", height(8)) xlabel(1989 1992 1995 1998 2001 2004 2007 2010 2013) ///
		graphregion(fcolor(white)) title("% of households holding any debt")
		graph export "$graphout\1.4 proportion of debtors.tif", as(tif) replace


*****some additional stuff
*check how debt to income ratios and actual debt holdings relate
	*who ware those with and without debt
		hist incperc [fw=wgtround] if debt>0 & debt!=. & imp==3
		hist incperc [fw=wgtround] if debt==0 & debt!=. & imp==3
	*do high debt to income ratio household contribute a lot to total borrowing
		*top 5% are having ratis above 4
			sum dy if year==2004 & imp==3,d
			total debt [fw=wgtround] if year==2004 & imp==3 //10.9 trillion
			total debt [fw=wgtround] if year==2004 & imp==3 & dy>4 & dy!=. //2.1 trillion; these are the highest 5% of dy
			total debt [fw=wgtround] if year==2004 & imp==3 & dy>4 & dy<10 & dy!=. //1.71 trillion
			total debt [fw=wgtround] if year==2004 & imp==3 & dy<0.43 & dy!=. //404 billion; these are the bottom 50%
			hist incperc [fw=wgtround] if year==2004 & imp==3 & dy<0.43 & dy!=.
			sum incper [fw=wgtround] if year==2004 & imp==3 & dy<0.43 & dy!=.,d //mean of 41 and median of 36
		*redo it for other year
			sum dy if year==1995 & imp==3,d
			total debt [fw=wgtround] if year==1995 & imp==3 //5.4 trillion
			total debt [fw=wgtround] if year==1995 & imp==3 & dy>3.5 & dy!=. //820 billion; these are the highest 5% of dy
			total debt [fw=wgtround] if year==1995 & imp==3 & dy<0.33 & dy!=. //220 billion; these are the bottom 50%
			hist incperc [fw=wgtround] if year==1995 & imp==3 & dy<0.433 & dy!=.
			sum incper [fw=wgtround] if year==1995 & imp==3 & dy<0.33 & dy!=.,d //mean of 42 and median of 38
	*interpretation
		*the highest indebted households by debt to income ratios are also hose which
		*contribute substantially to overal debt numbers.
		*The bottom 50% in terms of debt to income ratios only contribute less than 5% of total debt
		*while the top 5% of households in terms of debt to income ratios contribute abt. 20% of total debt.
	*formaliz and compare 1992 and 2004
		qui foreach year in 1992 2004 {
			mat share`year'=J(1,5,0)
			mat share50`year'=J(1,5,0)
```

```
	mat tot`year'=J(1,5,0)
	mat tot50`year'=J(1,5,0)
	forvalues i=1/5 {
		sum dy if year==`year' & imp==`i', d
		local cut=r(p95)
		local p75=r(p75)
		local p25=r(p25)
		total debt [fw=wgtround] if year==`year' & imp==`i'
		mat aux=e(b)
		total debt [fw=wgtround] if year==`year' & imp==`i' & dy>`cut' & dy!=.
		mat aux2=e(b)
		mat tot`year'[1,`i']=aux2[1,1]
		mat share`year'[1,`i']=aux2[1,1]/aux[1,1]
		total debt [fw=wgtround] if year==`year' & imp==`i' & dy>`p25' & dy<`p75' & dy!=.
		mat aux3=e(b)
		mat tot50`year'[1,`i']=aux3[1,1]
		mat share50`year'[1,`i']=aux3[1,1]/aux[1,1]
	}
	mat share`year'av=share`year'*J(5,1,0.2)
	mat share50`year'av=share50`year'*J(5,1,0.2)
	mat tot`year'av=tot`year'*J(5,1,0.2)
	mat tot50`year'av=tot50`year'*J(5,1,0.2)
	noi mat list share`year'av
	noi mat list share50`year'av
}
*
mat list share501992av
mat list tot1992av
mat list tot2004av
mat list tot501992av
mat list tot502004av
```

```
************************************************************************
*********************Picture 2: Debt Categories*********************
************************************************************************

***Debt components
	gen totmrthel=mrthel*wgt
	gen totinstall=install*wgt
	gen totresdbt=resdbt*wgt
	gen totccbal=ccbal*wgt
	gen totothloc=othloc*wgt
	gen totodebt=odebt*wgt
	*create aggregates
		foreach conc in mrthel install resdbt ccbal othloc odebt {
			forvalues year=1989(3)2013 {
				local im 5 //number of implicates needs to be specified
				quie forvalues imp=1/`im' {
					total tot`conc' if year==`year' & imp==`imp'
					mat tot`year'`conc'`imp' =(e(b))
					if `imp'==1 {
						local tot`year' "tot`year'`conc'`imp'"
					}
					else {
						local tot`year' "`tot`year''+tot`year'`conc'`imp'"
					}
				}
				disp "`tot`year''"
				qui mat tot`year'=(`tot`year'')/`im'
				qui mat colnames tot`year' = tot`conc'
				qui mat rownames tot`year' = `year'
			}
			*
			mat tot=(tot1989 \ tot1992 \ tot1995 \ tot1998 \ tot2001 \ tot2004 \ tot2007 \ tot2010 \ tot2013)
			mat list tot
			gen total`conc'=.
			forvalues i=1/9 {
				replace total`conc'=tot[`i',1]/10^9 in `i'
			}
		}
```

```
*plotting
	*as a share of total liabilities
		*produce stacked bar charts
			capture gen otherdebt=totalothloc+totalodebt+totalccbal
			label var totalmrthel "mortgages on primary residence"
			label var totalinstall "installment loans"
			label var totalresdbt "other mortgages"
			label var otherdebt "other debt"
			graph bar (asis) totalmrthel totalinstall totalresdbt otherdebt, over(yearplot) stack scheme(sj) ///
			legend(subtitle("")) ytit("% of total debt", height(8)) percentage ///
			graphregion(fcolor(white)) title("categories of household debt") ///
			bar(1, color(gs8)) bar(2, color(gs5)) bar(3, color(gs0)) bar(4, color(gs12))
			graph export "$graphout\2 debt categories1.tif", as(tif) replace
		*different split
			capture gen mortgages=totalmrthel+totalresdbt
			capture gen otherdebt2=totalothloc+totalodebt
			label var totalccbal "credit card debt"
			label var otherdebt2 "other debt"
			graph bar (asis) mortgages totalinstall totalccbal otherdebt2, over(yearplot) stack scheme(sj) ///
			legend(subtitle("")) ytit("% of total debt", height(8)) percentage ///
			graphregion(fcolor(white)) title("categories of household debt") ///
			bar(1, color(gs8)) bar(2, color(gs5)) bar(3, color(gs0)) bar(4, color(gs12))
			graph export "$graphout\3 debt categories2.tif", as(tif) replace
		*plot mortgages as a share
			capture gen prmortshare=totalmrthel/totaldebt*100
			capture gen resdbtshare=totalresdbt/totaldebt*100
			capture gen mortshare=mortgages/totaldebt*100
			*primary residence
				twoway line prmortshare yearplot, scheme(sj) ///
				legend(subtitle("")) ytit("% of total liabilities", height(8)) xlabel(1989 1992 1995 1998 2001 2004 2007 2010 2013) ///
				graphregion(fcolor(white)) title("mortages on primary residence")
				graph export "$graphout\3.1 primary residence mortgage.tif", as(tif) replace
			*other real estate related debt
				twoway line resdbtshare yearplot, scheme(sj) ///
				legend(subtitle("")) ytit("% of total liabilities", height(8)) xlabel(1989 1992 1995 1998 2001 2004 2007 2010 2013) ///
				graphregion(fcolor(white)) title("other real estate related debt")
				graph export "$graphout\3.2 other real estate debt.tif", as(tif) replace
			*both together
```

```
        twoway line mortshare yearplot, scheme(sj) ///
        legend(subtitle("")) ytit("% of total liabilities", height(8)) xlabel(1989 1992 1995 1998 2001 2004 2007 2010 2013) ///
        graphregion(fcolor(white)) title("real estate secured debt")
        graph export "$graphout\3.3 real estate secured debt.tif", as(tif) replace
*plot other components as a share
    capture gen installshare=totalinstall/totaldebt*100
    capture gen ccshare=totalccbal/totaldebt*100
    capture gen otherdebt2share=otherdebt2/totaldebt*100
*installment loans
    twoway line installshare yearplot, scheme(sj) ///
    legend(subtitle("")) ytit("% of total liabilities", height(8)) xlabel(1989 1992 1995 1998 2001 2004 2007 2010 2013) ///
    graphregion(fcolor(white)) title("instalment loans debt")
    graph export "$graphout\3.4 instalment loans.tif", as(tif) replace
*credit card balances
    twoway line ccshare yearplot, scheme(sj) ///
    legend(subtitle("")) ytit("% of total liabilities", height(8)) xlabel(1989 1992 1995 1998 2001 2004 2007 2010 2013) ///
    graphregion(fcolor(white)) title("credit card debt")
    graph export "$graphout\3.5 credit card debt.tif", as(tif) replace
*otherdebt 2 (other lines of credit and other debt)
    twoway line otherdebt2share yearplot, scheme(sj) ///
    legend(subtitle("")) ytit("% of total liabilities", height(8)) xlabel(1989 1992 1995 1998 2001 2004 2007 2010 2013) ///
    graphregion(fcolor(white)) title("other debt")
    graph export "$graphout\3.6 otherdebt2.tif", as(tif) replace


***********************************************************************
***********************Picture 3: Use of Debt***********************
***********************************************************************

***Debt components by use purpose
    *PLOAN1 Home purchase loans
    *PLOAN2 Home improvment loans
    *PLOAN3 vehicle loans
    *PLOAN4 loan for purchase of goods and services
    *PLOAN5 laons for investment
    *PLOAN6 loans for education
    *PLOAN7 mortgage loans for other real estate
```

```
	*PLOAN8 other unclassifiable loans
		forvalues i=1/8 {
			gen totploan`i'=PLOAN`i'*wgt
		}
	*create aggregates
		forvalues i=1/8 {
			forvalues year=1989(3)2013 {
				local im 5 //number of implicates needs to be specified
				quie forvalues imp=1/`im' {
					total totploan`i' if year==`year' & imp==`imp'
					mat totploan`i'`year'`imp' =(e(b))
					if `imp'==1 {
						local totploan`year' "totploan`i'`year'`imp'"
					}
					else {
						local totploan`year' "`totploan`year''+totploan`i'`year'`imp'"
					}
				}
				disp "`totploan`year''"
				qui mat totploan`year'=(`totploan`year'')/`im'
				qui mat colnames totploan`year' = totploan`i'
				qui mat rownames totploan`year' = `year'
			}
			*
			mat tot=(totploan1989 \ totploan1992 \ totploan1995 \ totploan1998 \ totploan2001 \ totploan2004 \ totploan2007 \ totploan2010 \ totploan2013)
			mat list tot
			gen totalploan`i'=.
			forvalues t=1/9 {
				replace totalploan`i'=tot[`t',1]/10^9 in `t'
			}
		}
*Plotting
	*relative to total liabilities
		*produce stack bar charts
			capture gen home=totalploan1+totalploan2
			capture gen consume=totalploan3+totalploan4
			capture gen invest=totalploan5+totalploan7
			label var home "home purch. and impr."
```

```
		label var consume "consumption"
		label var invest "investment"
		label var totalploan6 "education"
		graph bar (asis) home consume invest totalploan6, over(yearplot) stack scheme(sj) ///
		legend(subtitle("")) ytit("% of total debt", height(8)) percentage ///
		graphregion(fcolor(white)) title("use of household debt") ///
		bar(1, color(gs8)) bar(2, color(gs5)) bar(3, color(gs0)) bar(4, color(gs12))
		graph export "$graphout\3.7 use of debt.tif", as(tif) replace
	*plot as shares
		capture gen homeshare=home/totaldebt*100
		capture gen homepurchshare=totalploan1/totaldebt*100
		capture gen homeimpshare=totalploan2/totaldebt*100
		capture gen consushare=consume/totaldebt*100
		capture gen edushare=totalploan6/totaldebt*100
		*homepurchase and improvement
			twoway line homeshare yearplot, scheme(sj) ///
			legend(subtitle("")) ytit("% of total liabilities", height(8)) xlabel(1989 1992 1995 1998 2001 2004 2007 2010 2013) ///
			graphregion(fcolor(white)) title("home purchase and improvement")
			graph export "$graphout\3.8 home purch and improve share.tif", as(tif) replace
		*consumption
			twoway line consushare yearplot, scheme(sj) ///
			legend(subtitle("")) ytit("% of total liabilities", height(8)) xlabel(1989 1992 1995 1998 2001 2004 2007 2010 2013) ///
			graphregion(fcolor(white)) title("consumption")
			graph export "$graphout\3.9 consumption share.tif", as(tif) replace
		*education
			twoway line edushare yearplot, scheme(sj) ///
			legend(subtitle("")) ytit("% of total liabilities", height(8)) xlabel(1989 1992 1995 1998 2001 2004 2007 2010 2013) ///
			graphregion(fcolor(white)) title("education")
			graph export "$graphout\3.10 education share.tif", as(tif) replace
*relative to disposable income
	capture gen consincshare=consume/totalinc*100
```

```
************************************************************************
************Picture 4a: General Debt Inequality Nexus*****************
************************************************************************

**NOTE***
*Since for hispanic and other the distinction between college and non-college
*yields to small sample sizes, one can redo the excercise and define them only based on race.
*Similarly due to the size of the group of white households, there one could introduce more
*education categories, such as highschool!


*form edu X race groups
	gen eduaux=1 if edcl==4
	replace eduaux=0 if edcl<4
	sum eduaux edcl
	*race // 1=white-non-Hispanic 2=black 3=hispanic 5=other
	gen raceaux=1 if race==1
	replace raceaux=2 if race==2
	replace raceaux=3 if race==3
	gen raceaux2=1 if race==1
	replace raceaux2=2 if race==2
	count if raceaux==.
	count if race==.
*compute absolute top incomes and shares of top 1% and top 5%
*use education and race groupings
	gen top1redum=.
	gen top5redum=.
	foreach quant in 1 5 {
		mat av`quant'=J(30,9,1)
		mat share`quant'=J(30,9,1)
		mat obstot`quant'=J(30,9,1)
		mat obstop`quant'=J(30,9,1)
		mat weight`quant'=J(30,9,1)
		mat pop`quant'=J(30,9,1)
		mat debt`quant'=J(30,9,1)
	}
	local iter=0
	qui forvalues year=1989(3)2013 {
```

```
		nois disp "                                 `year'"
		local iter=`iter'+1
		local iter2=0
		forvalues i=1/5 {
			nois disp "          `i'"
			capture drop group
			egen group=group(eduaux raceaux) if year==`year' & imp==`i'
			forvalues j=1(1)6 { //specify number of groups
				local iter2=`iter2'+1
				foreach quant in 1 5 {
					local num=100-`quant'
					*the average measure
						_pctile income [pweight=wgt] if group==`j', p(`num')
						local cut=r(r1)
						sum income [fw=wgtround] if group==`j' & income>`cut'
						mat av`quant'[`iter2',`iter']=r(mean)
						replace top`quant'redum=1 if group==`j' & income>`cut'
					*the top income share measure
						total income [fw=wgtround] if group==`j'
						mat tot=e(b)
						total income [fw=wgtround] if group==`j' & income>`cut'
						mat share`quant'[`iter2',`iter']=e(b)/tot[1,1]
						replace top`quant'redum=1 if group==`j' & income>`cut'
					*total outstanding debt
						total debt [fw=wgtround] if group==`j' & income<=`cut'
						mat debt`quant'[`iter2',`iter']=e(b)
					*observations and mean weight for the groups
						sum wgt if group==`j' & income>`cut'
						mat weight`quant'[`iter2',`iter']=r(mean)
						mat obstop`quant'[`iter2',`iter']=r(N)
						sum income if group==`j'
						mat obstot`quant'[`iter2',`iter']=r(N)
						sum income [fw=wgtround] if group==`j' & income>`cut'
						mat pop`quant'[`iter2',`iter']=r(N)
				}
			}
		}
	}
```

```
	foreach var in av share obstot obstop weight debt pop {
		foreach quant in 1 5 {
			mat colnames `var'`quant'=1989 1992 1995 1998 2001 2004 2007 2010 2013
			mat rownames `var'`quant'=i1-nocol|w i1-nocol|b i1-nocol|h i1-col|w i1-col|b i1-col|h ///
									i2-nocol|w i2-nocol|b i2-nocol|h i2-col|w i2-col|b i2-col|h ///
									i3-nocol|w i3-nocol|b i3-nocol|h i3-col|w i3-col|b i3-col|h ///
									i4-nocol|w i4-nocol|b i4-nocol|h i4-col|w i4-col|b i4-col|h ///
									i5-nocol|w i5-nocol|b i5-nocol|h i5-col|w i5-col|b i5-col|h
		}
	}
*reorganize the matrices
	foreach var in av share obstot obstop weight debt pop {
		foreach quant in 1 5 {
			forvalues i=1/6 { //depends on numbers of groups!
				mat `var'`quant'aux`i'=(`var'`quant'[`i',1..9] \ `var'`quant'[`i'+6,1..9] \ `var'`quant'[`i'+12,1..9] \ `var'`quant'[`i'+18,1..9] \ `var'`quant'[`i'+24,1..9])
				if `i'==1 mat `var'`quant'new=`var'`quant'aux1
				if `i'>1 mat `var'`quant'new=(`var'`quant'new \ `var'`quant'aux`i')
			}
			mat list `var'`quant'new
		}
	}
*export the raw data per implicate
	cd "C:\Users\Rafael\Dropbox\Diss\paper 4\graphs"
	foreach var in av share obstot obstop weight debt pop {
		foreach quant in 1 5 {
			xml_tab `var'`quant'new, save("`var'`quant'.xml") replace
		}
	}
*run some checks
	_pctile income [fw=wgtround] if year==2004 & imp==3 & eduaux==1 & race==3, p(99)
	disp `r(r1)'
	total income [fw=wgtround] if year==2004 & imp==3 & eduaux==1 & race==3 & income>`r(r1)'
	mat test=e(b)
	total income [fw=wgtround] if year==2004 & imp==3 & eduaux==1 & race==3
	mat test2=e(b)
	mat test3=test[1,1]/test2[1,1]
	mat list test3
	mat list share1new
```

```
*average the data
	foreach var in  share debt {
		foreach quant in 1 5 {
			forvalues i=1/6 { //number of groups
				if `i'==1 mat `var'`quant'aux=(J(1,5,0.2),J(1,25,0))*`var'`quant'new
				if `i'==1 mat `var'`quant'final=`var'`quant'aux
				if (`i'>1 & `i'<6) mat  `var'`quant'aux=(J(1,5*(`i'-1),0),J(1,5,0.2),J(1,5*(6-`i'),0))*`var'`quant'new
				if `i'==6 mat  `var'`quant'aux=(J(1,25,0),J(1,5,0.2))*`var'`quant'new
				if `i'>1 mat `var'`quant'final=(`var'`quant'final \ `var'`quant'aux)
			}
			mat rownames `var'`quant'final=nocol|w nocol|b nocol|h col|w col|b col|h
			mat  list `var'`quant'final
		}
	}
*create variables
	local iter=0
	foreach group in nocolw nocolb nocolh colw colb colh {
		local iter=`iter'+1
		capture gen share1`group'=.
		capture gen share5`group'=.
		forvalues t=1/9 {
			replace share1`group'=share1final[`iter',`t']*100 in `t'
			replace share5`group'=share5final[`iter',`t']*100 in `t'
		}
	}
	local iter=0
	foreach group in nocolw nocolb nocolh colw colb colh {
		local iter=`iter'+1
		capture gen debt1`group'=.
		capture gen debt5`group'=.
		forvalues t=1/9 {
			replace debt1`group'=debt1final[`iter',`t']/10^9 in `t'
			replace debt5`group'=debt5final[`iter',`t']/10^9 in `t'
		}
	}
*plot
	*no college white
		twoway scatter share1nocolw debt1nocolw, scheme(sj) connect(l) ///
```

```
		sort(yearplot) mlabel(yearplot) mlabposition(9) legend(subtitle("")) title("no college, white") ///
		ytit("Top1% share", height(8)) xtit("outstanding liabilities, bn. 2013 US$", height(8)) graphregion(color(white))
		graph export "$graphout\4.1 nocol white.tif", as(tif) replace
	*no college black
		twoway scatter share1nocolb debt1nocolb, scheme(sj) connect(l) ///
		sort(yearplot) mlabel(yearplot) mlabposition(6) legend(subtitle("")) title("no college, black") ///
		ytit("Top1% share", height(8)) xtit("outstanding liabilities, bn. 2013 US$", height(8)) graphregion(color(white))
		graph export "$graphout\4.2 nocol black.tif", as(tif) replace
	*college white
		twoway scatter share1colw debt1colw, scheme(sj) connect(l) ///
		sort(yearplot) mlabel(yearplot) mlabposition(3) legend(subtitle("")) title("college, white") ///
		ytit("Top1% share", height(8)) xtit("outstanding liabilities, bn. 2013 US$", height(8)) graphregion(color(white))
		graph export "$graphout\4.3 col white.tif", as(tif) replace
	*college black
		twoway scatter share1colb debt1colb, scheme(sj) connect(l) ///
		sort(yearplot) mlabel(yearplot) mlabposition(4) legend(subtitle("")) title("college, black") ///
		ytit("Top1% share", height(8)) xtit("outstanding liabilities, bn. 2013 US$", height(8))	graphregion(color(white))
		graph export "$graphout\4.4 col black.tif", as(tif) replace
*summary plot
		label var debt1nocolw "no college, white"
		label var debt1nocolb "no college, black"
		label var debt1colw "college, white"
		label var debt1colb "college, black"
		graph bar (asis) debt1nocolw debt1colw debt1nocolb  debt1colb, over(yearplot) stack scheme(sj) ///
		legend(subtitle("")) ytit("share of group in %", height(8)) percentage ///
		graphregion(fcolor(white)) ///
		bar(1, color(gs8)) bar(2, color(gs0)) bar(3, color(gs5)) bar(4, color(gs12))
		graph export "$graphout\4.5 summary.tif", as(tif) replace
```

```
**********************************************************************
*************Picture 4b: Sustainability of borrowing*****************
**********************************************************************

*redo exercise of picture 4a but with debt to income ratios
*gen debt to income
	gen debtinc=debt/income
	*eliminate debt to income ratios above 100 because they are the result of very low
	*or implausibel low incomes
	replace debtinc=. if debtinc>100 //88 observations
	foreach quant in 1 5 {
		mat debtinc`quant'=J(30,9,1)
	}
	local iter=0
	qui forvalues year=1989(3)2013 {
		nois disp "                                        `year'"
		local iter=`iter'+1
		local iter2=0
		forvalues i=1/5 {
			nois disp "          `i'"
			capture drop group
			egen group=group(eduaux raceaux) if year==`year' & imp==`i'
			forvalues j=1(1)6 { //specify number of groups
				local iter2=`iter2'+1
				foreach quant in 1 5 {
					local num=100-`quant'
					*the average measure
						_pctile income [pweight=wgt] if group==`j', p(`num')
						local cut=r(r1)
					*total outstanding debt
						sum debtinc [fw=wgtround] if group==`j' & income<=`cut' ,d
						mat debtinc`quant'[`iter2',`iter']=r(p50)
				}
			}
		}
	}
	foreach var in debtinc {
		foreach quant in 1 5 {
```

```
				mat colnames `var'`quant'=1989 1992 1995 1998 2001 2004 2007 2010 2013
				mat rownames `var'`quant'=i1-nocol|w i1-nocol|b i1-nocol|h i1-col|w i1-col|b i1-col|h ///
							i2-nocol|w i2-nocol|b i2-nocol|h i2-col|w i2-col|b i2-col|h ///
							i3-nocol|w i3-nocol|b i3-nocol|h i3-col|w i3-col|b i3-col|h ///
							i4-nocol|w i4-nocol|b i4-nocol|h i4-col|w i4-col|b i4-col|h ///
							i5-nocol|w i5-nocol|b i5-nocol|h i5-col|w i5-col|b i5-col|h
			}
		}
*reorganize the matrices
		foreach var in debtinc {
			foreach quant in 1 5 {
				forvalues i=1/6 { //depends on numbers of groups!
					mat `var'`quant'aux`i'=(`var'`quant'[`i',1..9] \ `var'`quant'[`i'+6,1..9] \ `var'`quant'[`i'+12,1..9] \ `var'`quant'[`i'+18,1..9] \ `var'`quant'[`i'+24,1..9])
					if `i'==1 mat `var'`quant'new=`var'`quant'aux1
					if `i'>1 mat `var'`quant'new=(`var'`quant'new \ `var'`quant'aux`i')
				}
				mat list `var'`quant'new
			}
		}
*average the data
		foreach var in debtinc {
			foreach quant in 1 5 {
				forvalues i=1/6 { //number of groups
					if `i'==1 mat `var'`quant'aux=(J(1,5,0.2),J(1,25,0))*`var'`quant'new
					if `i'==1 mat `var'`quant'final=`var'`quant'aux
					if (`i'>1 & `i'<6) mat  `var'`quant'aux=(J(1,5*(`i'-1),0),J(1,5,0.2),J(1,5*(6-`i'),0))*`var'`quant'new
					if `i'==6 mat  `var'`quant'aux=(J(1,25,0),J(1,5,0.2))*`var'`quant'new
					if `i'>1 mat `var'`quant'final=(`var'`quant'final \ `var'`quant'aux)
				}
				mat rownames `var'`quant'final=nocol|w nocol|b nocol|h col|w col|b col|h
				mat  list `var'`quant'final
			}
		}
*create variables
		local iter=0
		foreach group in nocolw nocolb nocolh colw colb colh {
			local iter=`iter'+1
			capture gen debtinc1`group'=.
```

```
		capture gen debtinc5`group'=.
		forvalues t=1/9 {
			replace debtinc1`group'=debtinc1final[`iter',`t'] in `t'
			replace debtinc5`group'=debtinc5final[`iter',`t'] in `t'
		}
	}
*plot
	*no college white
		twoway scatter share1nocolw debtinc1nocolw, scheme(sj) connect(l) ///
		sort(yearplot) mlabel(yearplot) mlabposition(9) legend(subtitle("")) title("no college, white") ///
		ytit("Top1% share", height(8)) xtit("median debt to income ratio", height(8)) graphregion(color(white))
		graph export "$graphout\4b.1 median debt to income nocol white.tif", as(tif) replace
	*no college black
		twoway scatter share1nocolb debtinc1nocolb, scheme(sj) connect(l) ///
		sort(yearplot) mlabel(yearplot) mlabposition(6) legend(subtitle("")) title("no college, black") ///
		ytit("Top1% share", height(8)) xtit("median debt to income ratio", height(8)) graphregion(color(white))
		graph export "$graphout\4b.2 median debt to income nocol black.tif", as(tif) replace
	*college white
		twoway scatter share1colw debtinc1colw, scheme(sj) connect(l) ///
		sort(yearplot) mlabel(yearplot) mlabposition(3) legend(subtitle("")) title("college, white") ///
		ytit("Top1% share", height(8)) xtit("median debt to income ratio", height(8)) graphregion(color(white))
		graph export "$graphout\4b.3 median debt to income col white.tif", as(tif) replace
	*college black
		twoway scatter share1colb debtinc1colb, scheme(sj) connect(l) ///
		sort(yearplot) mlabel(yearplot) mlabposition(1) legend(subtitle("")) title("college, black") ///
		ytit("Top1% share", height(8)) xtit("median debt to income ratio", height(8))	graphregion(color(white))
		graph export "$graphout\4b.4 median debt to income col black.tif", as(tif) replace
*top 5
	*no college white
		twoway scatter share5nocolw debtinc1nocolw, scheme(sj) connect(l) ///
		sort(yearplot) mlabel(yearplot) mlabposition(9) legend(subtitle("")) title("no college, white") ///
		ytit("Top5% share", height(8)) xtit("median debt to income ratio", height(8)) graphregion(color(white))
		graph export "$graphout\4b.1 median debt to income nocol white top5.tif", as(tif) replace
	*no college black
		twoway scatter share5nocolb debtinc1nocolb, scheme(sj) connect(l) ///
		sort(yearplot) mlabel(yearplot) mlabposition(6) legend(subtitle("")) title("no college, black") ///
		ytit("Top5% share", height(8)) xtit("median debt to income ratio", height(8)) graphregion(color(white))
		graph export "$graphout\4b.2 median debt to income nocol black top5.tif", as(tif) replace
```

```
    *college white
        twoway scatter share5colw debtinc1colw, scheme(sj) connect(l) ///
        sort(yearplot) mlabel(yearplot) mlabposition(3) legend(subtitle("")) title("college, white") ///
        ytit("Top5% share", height(8)) xtit("median debt to income ratio", height(8)) graphregion(color(white))
        graph export "$graphout\4b.3 median debt to income col white top5.tif", as(tif) replace
    *college black
        twoway scatter share5colb debtinc1colb, scheme(sj) connect(l) ///
        sort(yearplot) mlabel(yearplot) mlabposition(1) legend(subtitle("")) title("college, black") ///
        ytit("Top5% share", height(8)) xtit("median debt to income ratio", height(8))    graphregion(color(white))
        graph export "$graphout\4b.4 median debt to income col black top5.tif", as(tif) replace



**********************************************************************
*****Picture 4c: Proportion of credit constrained households**********
**********************************************************************
    sum X407
    *In the past five years, has a particular lender or creditor
  *turned down any request you (or your {husband/wife/partner}) made for credit, or not given you
  *as much credit as you applied for?
    *(1=yes turned down, 3=yes not as much, 5=No, 0=inappr)
    sum turndown
    *turndown=1 if X407==1
    forvalues year=1989(3)2013 {
        local im 5 //number of implicates needs to be specified
        qui forvalues imp=1/`im' {
            total wgt if year==`year' & imp==`imp' & debt>0 & debt!=.
            mat totpop=e(b)
            total wgt if year==`year' & imp==`imp' & turndown==1 & debt!=.
            mat turndown`year'`imp'=e(b)/totpop[1,1]
            if `imp'==1 {
                local turndown`year' "turndown`year'`imp'"
            }
            else {
                local turndown`year' "`turndown`year''+turndown`year'`imp'"
            }
        }
        disp "`turndown`year''"
```

```
		qui mat turndown`year'=(`turndown`year'')/`im'
		qui mat colnames turndown`year' = turndown
		qui mat rownames turndown`year' = `year'
	}
	*
	mat turndown=(turndown1989 \ turndown1992 \ turndown1995 \ turndown1998 \ turndown2001 \ turndown2004 \ turndown2007 \ turndown2010 \ turndown2013)
	mat list turndown
	capture drop turndownshare
	gen turndownshare=.
	forvalues i=1/9 {
		replace turndownshare=turndown[`i',1]*100 in `i'
	}
	*plot aggregate
		twoway line turndownshare yearplot, scheme(sj) ///
		legend(subtitle("")) ytit("% of indebted households", height(8)) xlabel(1989 1992 1995 1998 2001 2004 2007 2010 2013) ///
		graphregion(fcolor(white)) title("denied borrowing")
		graph export "$graphout\1.6 proportion of turned down households.tif", as(tif) replace

*compute share per group
	foreach quant in 1 5 {
		mat turndown`quant'=J(30,9,1)
	}
	local iter=0
	qui forvalues year=1989(3)2013 {
		nois disp "				`year'"
		local iter=`iter'+1
		local iter2=0
		forvalues i=1/5 {
			nois disp "		`i'"
			capture drop group
			egen group=group(eduaux raceaux) if year==`year' & imp==`i'
			forvalues j=1(1)6 { //specify number of groups
				local iter2=`iter2'+1
				foreach quant in 1 { //for top1% only or top5% as well
					local num=100-`quant'
					*the cut-off
						_pctile income [pweight=wgt] if group==`j', p(`num')
						local cut=r(r1)
```

```
					*turned down households
						total wgt if group==`j' & income<=`cut' & debt>0 & debt!=.
						mat totpop=e(b)
						total wgt if group==`j' & income<=`cut' & turndown==1 & debt!=.
						mat turndown`quant'[`iter2',`iter']=e(b)/totpop[1,1]
				}
			}
		}
	}
	foreach var in turndown {
		foreach quant in 1 { // top1% only or top5% as well
			mat colnames `var'`quant'=1989 1992 1995 1998 2001 2004 2007 2010 2013
			mat rownames `var'`quant'=i1-nocol|w i1-nocol|b i1-nocol|h i1-col|w i1-col|b i1-col|h ///
										i2-nocol|w i2-nocol|b i2-nocol|h i2-col|w i2-col|b i2-col|h ///
										i3-nocol|w i3-nocol|b i3-nocol|h i3-col|w i3-col|b i3-col|h ///
										i4-nocol|w i4-nocol|b i4-nocol|h i4-col|w i4-col|b i4-col|h ///
										i5-nocol|w i5-nocol|b i5-nocol|h i5-col|w i5-col|b i5-col|h
		}
	}
*reorganize the matrices
	foreach var in turndown {
		foreach quant in 1 { // top1% only or top5% as well
			forvalues i=1/6 { //depends on numbers of groups!
				mat `var'`quant'aux`i'=(`var'`quant'[`i',1..9] \ `var'`quant'[`i'+6,1..9] \ `var'`quant'[`i'+12,1..9] \ `var'`quant'[`i'+18,1..9] \ `var'`quant'[`i'+24,1..9])
				if `i'==1 mat `var'`quant'new=`var'`quant'aux1
				if `i'>1 mat `var'`quant'new=(`var'`quant'new \ `var'`quant'aux`i')
			}
			mat list `var'`quant'new
		}
	}
*average the data
	foreach var in turndown {
		foreach quant in 1 { // top1% only or top5% as well
			forvalues i=1/6 { //number of groups
				if `i'==1 mat `var'`quant'aux=(J(1,5,0.2),J(1,25,0))*`var'`quant'new
				if `i'==1 mat `var'`quant'final=`var'`quant'aux
				if (`i'>1 & `i'<6) mat  `var'`quant'aux=(J(1,5*(`i'-1),0),J(1,5,0.2),J(1,5*(6-`i'),0))*`var'`quant'new
				if `i'==6 mat  `var'`quant'aux=(J(1,25,0),J(1,5,0.2))*`var'`quant'new
```

```
                        if `i'>1 mat `var'`quant'final=(`var'`quant'final \ `var'`quant'aux)
                }
                mat rownames `var'`quant'final=nocol|w nocol|b nocol|h col|w col|b col|h
                mat  list `var'`quant'final
        }
    }
*create variables
    local iter=0
    foreach group in nocolw nocolb nocolh colw colb colh {
        local iter=`iter'+1
        capture gen turndown`group'=.
        forvalues t=1/9 {
                replace turndown`group'=turndown1final[`iter',`t']*100 in `t'
        }
    }
*plot
    *no college white
        twoway scatter turndownnocolw yearplot, scheme(sj) connect(l) ///
        sort(yearplot) mlabel(yearplot) mlabposition(12) legend(subtitle("")) title("no college, white") ///
        ytit("% of households turned down", height(8)) xtit("", height(8)) graphregion(color(white))
        graph export "$graphout\4c.1 turndown nocol white.tif", as(tif) replace
    *no college black
        twoway scatter turndownnocolb yearplot, scheme(sj) connect(l) ///
        sort(yearplot) mlabel(yearplot) mlabposition(12) legend(subtitle("")) title("no college, black") ///
        ytit("% of households turned down", height(8)) xtit("", height(8)) graphregion(color(white))
        graph export "$graphout\4c.2 turndown nocol black.tif", as(tif) replace
    *college white
        twoway scatter turndowncolw yearplot, scheme(sj) connect(l) ///
        sort(yearplot) mlabel(yearplot) mlabposition(12) legend(subtitle("")) title("college, white") ///
        ytit("% of households turned down", height(8)) xtit("", height(8)) graphregion(color(white))
        graph export "$graphout\4c.3 turndown col white.tif", as(tif) replace
    *college black
        twoway scatter turndowncolb yearplot, scheme(sj) connect(l) ///
        sort(yearplot) mlabel(yearplot) mlabposition(12) legend(subtitle("")) title("college, black") ///
        ytit("% of households turned down", height(8)) xtit("", height(8))       graphregion(color(white))
        graph export "$graphout\4c.4 turndown col black.tif", as(tif) replace
```

```
**********************************************************************
*************Picture 4d: Development of housing wealth****************
**********************************************************************

*generate residential real estate variable
	gen resre=houses+oresre
*compute houisng wealth per group
	foreach quant in 1 {
		mat resre`quant'=J(30,9,1)
		mat rep`quant'=J(30,9,1)
	}
	local iter=0
	qui forvalues year=1989(3)2013 {
		nois disp "						`year'"
		local iter=`iter'+1
		local iter2=0
		forvalues i=1/5 {
			nois disp "		`i'"
			capture drop group
			egen group=group(eduaux raceaux) if year==`year' & imp==`i'
			forvalues j=1(1)6 { //specify number of groups
				local iter2=`iter2'+1
				foreach quant in 1 { //for top1% only or top5% as well
					local num=100-`quant'
					*the cut-off
						_pctile income [pweight=wgt] if group==`j', p(`num')
						local cut=r(r1)
					*housing wealth and purchases
						total resre [pweight=wgt] if group==`j' & income<=`cut'
						mat resre`quant'[`iter2',`iter']=e(b)
						total REP [pweight=wgt] if group==`j' & income<=`cut'
						mat rep`quant'[`iter2',`iter']=e(b)
				}
			}
		}
	}
	foreach var in resre rep {
		foreach quant in 1 { // top1% only or top5% as well
```

```
				mat colnames `var'`quant'=1989 1992 1995 1998 2001 2004 2007 2010 2013
				mat rownames `var'`quant'=i1-nocol|w i1-nocol|b i1-nocol|h i1-col|w i1-col|b i1-col|h ///
											i2-nocol|w i2-nocol|b i2-nocol|h i2-col|w i2-col|b i2-col|h ///
											i3-nocol|w i3-nocol|b i3-nocol|h i3-col|w i3-col|b i3-col|h ///
											i4-nocol|w i4-nocol|b i4-nocol|h i4-col|w i4-col|b i4-col|h ///
											i5-nocol|w i5-nocol|b i5-nocol|h i5-col|w i5-col|b i5-col|h
			}
		}
*reorganize the matrices
		foreach var in resre rep {
			foreach quant in 1 { // top1% only or top5% as well
				forvalues i=1/6 { //depends on numbers of groups!
					mat `var'`quant'aux`i'=(`var'`quant'[`i',1..9] \ `var'`quant'[`i'+6,1..9] \ `var'`quant'[`i'+12,1..9] \ `var'`quant'[`i'+18,1..9] \ `var'`quant'[`i'+24,1..9])
					if `i'==1 mat `var'`quant'new=`var'`quant'aux1
					if `i'>1 mat `var'`quant'new=(`var'`quant'new \ `var'`quant'aux`i')
				}
				mat list `var'`quant'new
			}
		}
*average the data
		foreach var in resre rep {
			foreach quant in 1 { // top1% only or top5% as well
				forvalues i=1/6 { //number of groups
					if `i'==1 mat `var'`quant'aux=(J(1,5,0.2),J(1,25,0))*`var'`quant'new
					if `i'==1 mat `var'`quant'final=`var'`quant'aux
					if (`i'>1 & `i'<6) mat  `var'`quant'aux=(J(1,5*(`i'-1),0),J(1,5,0.2),J(1,5*(6-`i'),0))*`var'`quant'new
					if `i'==6 mat  `var'`quant'aux=(J(1,25,0),J(1,5,0.2))*`var'`quant'new
					if `i'>1 mat `var'`quant'final=(`var'`quant'final \ `var'`quant'aux)
				}
				mat rownames `var'`quant'final=nocol|w nocol|b nocol|h col|w col|b col|h
				mat  list `var'`quant'final
			}
		}
*create variables
		local iter=0
		foreach group in nocolw nocolb nocolh colw colb colh {
			local iter=`iter'+1
			capture gen resre`group'=.
```

```
			capture gen rep`group'=.
			forvalues t=1/9 {
				replace resre`group'=resre1final[`iter',`t']/10^9 in `t'
				replace rep`group'=(rep1final[`iter',`t']/10^9)*3 in `t'
			}
		}
*plot resre
		*no college white
			twoway scatter resrenocolw debt1nocolw, scheme(sj) connect(l) ///
			sort(yearplot) mlabel(yearplot) mlabposition(12) legend(subtitle("")) title("no college, white") ///
			ytit("housing wealth, bn. 2013 US$", height(8)) xtit("outstanding liabilities, bn. 2013 US$", height(8)) ///
			graphregion(color(white))
			graph export "$graphout\4d.1 resre nocol white.tif", as(tif) replace
		*no college black
			twoway scatter resrenocolb debt1nocolb, scheme(sj) connect(l) ///
			sort(yearplot) mlabel(yearplot) mlabposition(12) legend(subtitle("")) title("no college, black") ///
			ytit("housing wealth, bn. 2013 US$", height(8)) xtit("outstanding liabilities, bn. 2013 US$", height(8)) ///
			graphregion(color(white))
			graph export "$graphout\4d.2 resre nocol black.tif", as(tif) replace
		*college white
			twoway scatter resrecolw debt1colw, scheme(sj) connect(l) ///
			sort(yearplot) mlabel(yearplot) mlabposition(12) legend(subtitle("")) title("college, white") ///
			ytit("housing wealth, bn. 2013 US$", height(8)) xtit("outstanding liabilities, bn. 2013 US$", height(8)) ///
			graphregion(color(white))
			graph export "$graphout\4d.3 resre col white.tif", as(tif) replace
		*college black
			twoway scatter resrecolb debt1colb, scheme(sj) connect(l) ///
			sort(yearplot) mlabel(yearplot) mlabposition(12) legend(subtitle("")) title("college, black") ///
			ytit("housing wealth, bn. 2013 US$", height(8)) xtit("outstanding liabilities, bn. 2013 US$", height(8))		///
			graphregion(color(white))
			graph export "$graphout\4d.4 resre col black.tif", as(tif) replace
*plot resre		(with median instead of absolute values)
		*no college white
			twoway scatter resrenocolw debtinc1nocolw, scheme(sj) connect(l) ///
			sort(yearplot) mlabel(yearplot) mlabposition(12) legend(subtitle("")) title("no college, white") ///
			ytit("housing wealth, bn. 2013 US$", height(8)) xtit("median debt to income ratio", height(8)) ///
			graphregion(color(white))
			graph export "$graphout\4d.1 resre nocol white median.tif", as(tif) replace
```

```
	*no college black
		twoway scatter resrenocolb debtinc1nocolb, scheme(sj) connect(l) ///
		sort(yearplot) mlabel(yearplot) mlabposition(12) legend(subtitle("")) title("no college, black") ///
		ytit("housing wealth, bn. 2013 US$", height(8)) xtit("median debt to income ratio", height(8)) ///
		graphregion(color(white))
		graph export "$graphout\4d.2 resre nocol black median.tif", as(tif) replace
	*college white
		twoway scatter resrecolw debtinc1colw, scheme(sj) connect(l) ///
		sort(yearplot) mlabel(yearplot) mlabposition(12) legend(subtitle("")) title("college, white") ///
		ytit("housing wealth, bn. 2013 US$", height(8)) xtit("median debt to income ratio", height(8)) ///
		graphregion(color(white))
		graph export "$graphout\4d.3 resre col white median.tif", as(tif) replace
	*college black
		twoway scatter resrecolb debtinc1colb, scheme(sj) connect(l) ///
		sort(yearplot) mlabel(yearplot) mlabposition(12) legend(subtitle("")) title("college, black") ///
		ytit("housing wealth, bn. 2013 US$", height(8)) xtit("median debt to income ratio", height(8))	///
		graphregion(color(white))
		graph export "$graphout\4d.4 resre col black median.tif", as(tif) replace
*plot REP
	*gen differenced debt data
		foreach var in debt1nocolw debt1nocolb debt1colw debt1colb {
			capture drop `var'dif
			gen `var'dif=`var'[_n]-`var'[_n-1]
		}
	*no college white
		twoway (scatter repnocolw debt1nocolwdif, sort(yearplot) mlabel(yearplot) mlabposition(12)) ///
		|| (function y=x, range(0 1000)) , scheme(sj) ///
		legend(off) title("no college, white") ///
		ytit("real estate purchases current year times 3", height(8)) ///
		xtit("change to previous observation in outstanding liabilities", height(8)) ///
		graphregion(color(white))
		graph export "$graphout\4d.1 rep nocol white.tif", as(tif) replace
	*no college black
		twoway (scatter repnocolb debt1nocolbdif, sort(yearplot) mlabel(yearplot) mlabposition(12)) ///
		(function y=x, range(0 150)), scheme(sj)  ///
		legend(off) title("no college, black") ///
		ytit("real estate purchases current year times 3", height(8)) ///
		xtit("change to previous observation in outstanding liabilities", height(8)) ///
```

```
		graphregion(color(white))
		graph export "$graphout\4d.2 rep nocol black.tif", as(tif) replace
	*college white
		twoway (scatter repcolw debt1colwdif, sort(yearplot) mlabel(yearplot) mlabposition(12)) ///
		|| (function y=x, range(0 1500)), scheme(sj)  ///
		legend(off) title("college, white") ///
		ytit("real estate purchases current year times 3", height(8)) ///
		xtit("change to previous observation in outstanding liabilities", height(8)) ///
		graphregion(color(white))
		graph export "$graphout\4d.3 rep col white.tif", as(tif) replace
	*college black
		twoway (scatter repcolb debt1colbdif, sort(yearplot) mlabel(yearplot) mlabposition(12)) ///
		|| (function y=x, range(0 150)), scheme(sj)  ///
		legend(off) title("college, black") ///
		ytit("real estate purchases current year times 3", height(8)) ///
		xtit("change to previous observation in outstanding liabilities", height(8))		///
		graphregion(color(white))
		graph export "$graphout\4d.4 rep col black.tif", as(tif) replace




**********************************************************************
**************Picture 5: The role of homeownership********************
**********************************************************************

*compute absolute top incomes and shares of top 1% and top 5%
*use education and race groupings
	foreach quant in 1 5 {
		mat debthome`quant'=J(30,9,1)
	}
	*housecl=1: owner of primary residence
	*X606 year purchased the site (rents home, owns site)
	*X616 year purchased mobile home (owns home, rents site)
	*X626 year purchased mobile home (owns home and site)
	*X630 year purchased site (owns home and site)
```

```
*X634 year purchased home and site (owns home and site)
*X720 year purchased any part (owns part of home)
local iter=0
qui forvalues year=1989(3)2013 {
	nois disp "					`year'"
	local iter=`iter'+1
	local iter2=0
	forvalues i=1/5 {
		nois disp "		`i'"
		capture drop group
		egen group=group(eduaux raceaux) if year==`year' & imp==`i'
		forvalues j=1(1)6 { //specify number of groups
			local iter2=`iter2'+1
			foreach quant in 1 5 {
				local num=100-`quant'
				*the average measure
					_pctile income [pweight=wgt] if group==`j', p(`num')
					local cut=r(r1)
				*total outstanding debt
					total debt [fw=wgtround] if group==`j' & income<=`cut' & housecl==1 & ///
					X606!=`year' & X616!=`year' & X626!=`year' & X630!=`year' & X634!=`year' & X720!=`year'
					mat debthome`quant'[`iter2',`iter']=e(b)
			}
		}
	}
}
foreach var in debthome {
	foreach quant in 1 5 {
		mat colnames `var'`quant'=1989 1992 1995 1998 2001 2004 2007 2010 2013
		mat rownames `var'`quant'=i1-nocol|w i1-nocol|b i1-nocol|h i1-col|w i1-col|b i1-col|h ///
							i2-nocol|w i2-nocol|b i2-nocol|h i2-col|w i2-col|b i2-col|h ///
							i3-nocol|w i3-nocol|b i3-nocol|h i3-col|w i3-col|b i3-col|h ///
							i4-nocol|w i4-nocol|b i4-nocol|h i4-col|w i4-col|b i4-col|h ///
							i5-nocol|w i5-nocol|b i5-nocol|h i5-col|w i5-col|b i5-col|h
	}
}
*reorganize the matrices
foreach var in debthome {
```

```
		foreach quant in 1 5 {
			forvalues i=1/6 { //depends on numbers of groups!
				mat `var'`quant'aux`i'=(`var'`quant'[`i',1..9] \ `var'`quant'[`i'+6,1..9] \ `var'`quant'[`i'+12,1..9] \ `var'`quant'[`i'+18,1..9] \ `var'`quant'[`i'+24,1..9])
				if `i'==1 mat `var'`quant'new=`var'`quant'aux1
				if `i'>1 mat `var'`quant'new=(`var'`quant'new \ `var'`quant'aux`i')
			}
			mat list `var'`quant'new
		}
	}
*average the data
	foreach var in debthome {
		foreach quant in 1 5 {
			forvalues i=1/6 { //number of groups
				if `i'==1 mat `var'`quant'aux=(J(1,5,0.2),J(1,25,0))*`var'`quant'new
				if `i'==1 mat `var'`quant'final=`var'`quant'aux
				if (`i'>1 & `i'<6) mat  `var'`quant'aux=(J(1,5*(`i'-1),0),J(1,5,0.2),J(1,5*(6-`i'),0))*`var'`quant'new
				if `i'==6 mat  `var'`quant'aux=(J(1,25,0),J(1,5,0.2))*`var'`quant'new
				if `i'>1 mat `var'`quant'final=(`var'`quant'final \ `var'`quant'aux)
			}
			mat rownames `var'`quant'final=nocol|w nocol|b nocol|h col|w col|b col|h
			mat  list `var'`quant'final
		}
	}
*create variables
	local iter=0
	foreach group in nocolw nocolb nocolh colw colb colh {
		local iter=`iter'+1
		capture gen debthome1`group'=.
		capture gen debthome5`group'=.
		forvalues t=1/9 {
			replace debthome1`group'=debthome1final[`iter',`t']/10^9 in `t'
			replace debthome5`group'=debthome5final[`iter',`t']/10^9 in `t'
		}
	}
	local iter=0
	foreach group in nocolw nocolb nocolh colw colb colh {
		local iter=`iter'+1
		capture gen debtnohome1`group'=.
```

```
			capture gen debtnohome5`group'=.
			forvalues t=1/9 {
				replace debtnohome1`group'=debt1final[`iter',`t']/10^9-debthome1final[`iter',`t']/10^9 in `t'
				replace debtnohome5`group'=debt5final[`iter',`t']/10^9-debthome5final[`iter',`t']/10^9 in `t'
			}
		}
*plot
		*no college white
			twoway scatter share1nocolw debtnohome1nocolw, scheme(sj) connect(l) ///
			sort(yearplot) mlabel(yearplot) mlabposition(9) legend(subtitle("")) title("non-home-owners, no college, white") ///
			ytit("Top1% share", height(8)) xtit("outstanding liabilities, bn. 2013 US$", height(8)) graphregion(color(white))
			graph export "$graphout\5.1 nohome nocol white.tif", as(tif) replace
		*no college black
			twoway scatter share1nocolb debtnohome1nocolb, scheme(sj) connect(l) ///
			sort(yearplot) mlabel(yearplot) mlabposition(6) legend(subtitle("")) title("non-home-owners, no college, black") ///
			ytit("Top1% share", height(8)) xtit("outstanding liabilities, bn. 2013 US$", height(8)) graphregion(color(white))
			graph export "$graphout\5.2 nohome nocol black.tif", as(tif) replace
		*college white
			twoway scatter share1colw debtnohome1colw, scheme(sj) connect(l) ///
			sort(yearplot) mlabel(yearplot) mlabposition(3) legend(subtitle("")) title("non-home-owners, college, white") ///
			ytit("Top1% share", height(8)) xtit("outstanding liabilities, bn. 2013 US$", height(8)) graphregion(color(white))
			graph export "$graphout\5.3 nohome col white.tif", as(tif) replace
		*college black
			twoway scatter share1colb debtnohome1colb, scheme(sj) connect(l) ///
			sort(yearplot) mlabel(yearplot) mlabposition(4) legend(subtitle("")) title("non-home-owners, college, black") ///
			ytit("Top1% share", height(8)) xtit("outstanding liabilities, bn. 2013 US$", height(8))	graphregion(color(white))
			graph export "$graphout\5.4 nohome col black.tif", as(tif) replace
*summary plot
			label var debtnohome1nocolw "no college, white"
			label var debtnohome1nocolb "no college, black"
			label var debtnohome1colw "college, white"
			label var debtnohome1colb "college, black"
			graph bar (asis) debtnohome1nocolw debtnohome1colw debtnohome1nocolb  debtnohome1colb, over(yearplot) stack scheme(sj) ///
			legend(subtitle("")) ytit("share of group in %", height(8)) percentage ///
			graphregion(fcolor(white)) title("borrowing non-homeowners by groups") ///
			bar(1, color(gs8)) bar(2, color(gs0)) bar(3, color(gs5)) bar(4, color(gs12))
			graph export "$graphout\5.5 summary.tif", as(tif) replace
```

```
*******************************************************************
***Picture 6: Linking Consumption and Education Loans to Inequality***
*******************************************************************

*compute absolute top incomes and shares of top 1% and top 5%
*use education and race groupings
	foreach quant in 1 5 {
		mat consedu`quant'=J(30,9,1)
	}
	capture drop consedu
	gen consedu=PLOAN3+PLOAN4+PLOAN6 // loans for consumption of goods and services, cars and education
	local iter=0
	qui forvalues year=1989(3)2013 {
		nois disp "					`year'"
		local iter=`iter'+1
		local iter2=0
		forvalues i=1/5 {
			nois disp "		`i'"
			capture drop group
			egen group=group(eduaux raceaux) if year==`year' & imp==`i'
			forvalues j=1(1)6 { //specify number of groups
				local iter2=`iter2'+1
				foreach quant in 1 5 {
					local num=100-`quant'
					*the average measure
						_pctile income [pweight=wgt] if group==`j', p(`num')
						local cut=r(r1)
					*total outstanding debt
						total consedu [fw=wgtround] if group==`j' & income<=`cut'
						mat consedu`quant'[`iter2',`iter']=e(b)
				}
			}
		}
	}
	foreach var in consedu {
		foreach quant in 1 5 {
			mat colnames `var'`quant'=1989 1992 1995 1998 2001 2004 2007 2010 2013
			mat rownames `var'`quant'=i1-nocol|w i1-nocol|b i1-nocol|h i1-col|w i1-col|b i1-col|h ///
```

```
                                                                                    i2-nocol|w i2-nocol|b i2-nocol|h i2-col|w i2-col|b i2-col|h ///
                                                                                    i3-nocol|w i3-nocol|b i3-nocol|h i3-col|w i3-col|b i3-col|h ///
                                                                                    i4-nocol|w i4-nocol|b i4-nocol|h i4-col|w i4-col|b i4-col|h ///
                                                                                    i5-nocol|w i5-nocol|b i5-nocol|h i5-col|w i5-col|b i5-col|h
                }
        }
*reorganize the matrices
        foreach var in consedu {
                foreach quant in 1 5 {
                        forvalues i=1/6 { //depends on numbers of groups!
                                mat `var'`quant'aux`i'=(`var'`quant'[`i',1..9] \ `var'`quant'[`i'+6,1..9] \ `var'`quant'[`i'+12,1..9] \ `var'`quant'[`i'+18,1..9] \ `var'`quant'[`i'+24,1..9])
                                if `i'==1 mat `var'`quant'new=`var'`quant'aux1
                                if `i'>1 mat `var'`quant'new=(`var'`quant'new \ `var'`quant'aux`i')
                        }
                        mat list `var'`quant'new
                }
        }
*average the data
        foreach var in consedu {
                foreach quant in 1 5 {
                        forvalues i=1/6 { //number of groups
                                if `i'==1 mat `var'`quant'aux=(J(1,5,0.2),J(1,25,0))*`var'`quant'new
                                if `i'==1 mat `var'`quant'final=`var'`quant'aux
                                if (`i'>1 & `i'<6) mat  `var'`quant'aux=(J(1,5*(`i'-1),0),J(1,5,0.2),J(1,5*(6-`i'),0))*`var'`quant'new
                                if `i'==6 mat  `var'`quant'aux=(J(1,25,0),J(1,5,0.2))*`var'`quant'new
                                if `i'>1 mat `var'`quant'final=(`var'`quant'final \ `var'`quant'aux)
                        }
                        mat rownames `var'`quant'final=nocol|w nocol|b nocol|h col|w col|b col|h
                        mat  list `var'`quant'final
                }
        }
*create variables
        local iter=0
        foreach group in nocolw nocolb nocolh colw colb colh {
                local iter=`iter'+1
                capture gen consedu1`group'=.
                capture gen consedu5`group'=.
                forvalues t=1/9 {
```

```
				replace consedu1`group'=consedu1final[`iter',`t']/10^9 in `t'
				replace consedu5`group'=consedu5final[`iter',`t']/10^9 in `t'
			}
		}
*plot
		*no college white
			twoway scatter share1nocolw consedu1nocolw, scheme(sj) connect(l) ///
			sort(yearplot) mlabel(yearplot) mlabposition(9) legend(subtitle("")) title("no college, white") ///
			ytit("Top1% share", height(8)) xtit("outstanding consumption+education liabilities, bn. 2013 US$", height(8)) graphregion(color(white))
			graph export "$graphout\6.1 consedu nocol white.tif", as(tif) replace
		*no college black
			twoway scatter share1nocolb consedu1nocolb, scheme(sj) connect(l) ///
			sort(yearplot) mlabel(yearplot) mlabposition(6) legend(subtitle("")) title("no college, black") ///
			ytit("Top1% share", height(8)) xtit("outstanding consumption+education liabilities, bn. 2013 US$", height(8)) graphregion(color(white))
			graph export "$graphout\6.2 consedu nocol black.tif", as(tif) replace
		*college white
			twoway scatter share1colw consedu1colw, scheme(sj) connect(l) ///
			sort(yearplot) mlabel(yearplot) mlabposition(3) legend(subtitle("")) title("college, white") ///
			ytit("Top1% share", height(8)) xtit("outstanding consumption+education liabilities, bn. 2013 US$", height(8)) graphregion(color(white))
			graph export "$graphout\6.3 consedu col white.tif", as(tif) replace
		*college black
			twoway scatter share1colb consedu1colb, scheme(sj) connect(l) ///
			sort(yearplot) mlabel(yearplot) mlabposition(4) legend(subtitle("")) title("college, black") ///
			ytit("Top1% share", height(8)) xtit("outstanding consumption+education liabilities, bn. 2013 US$", height(8))graphregion(color(white))
			graph export "$graphout\6.4 consedu col black.tif", as(tif) replace
*summary plot
			label var consedu1nocolw "no college, white"
			label var consedu1nocolb "no college, black"
			label var consedu1colw "college, white"
			label var consedu1colb "college, black"
			graph bar (asis) consedu1nocolw consedu1colw consedu1nocolb  consedu1colb, over(yearplot) stack scheme(sj) ///
			legend(subtitle("")) ytit("share of group in %", height(8)) percentage ///
			graphregion(fcolor(white)) title("borrowing for consumption and education by groups") ///
			bar(1, color(gs8)) bar(2, color(gs0)) bar(3, color(gs5)) bar(4, color(gs12))
			graph export "$graphout\6.5 summary.tif", as(tif) replace
```

```
**********************************************************************
******Picture 7: Linking Borrowing and Negative Income Shocks*********
**********************************************************************

*define low income dummies
	gen incrat=income/norminc
	sum incrat,d
	gen lowincd=0
	replace lowincd=1 if incrat<1 & incrat!=.
***Aggregate Total Debt********************************************
	gen totdebtlowinc=debt*wgt*lowincd
	forvalues year=1989(3)2013 {
		local im 5 //number of implicates needs to be specified
		quie forvalues imp=1/`im' {
			total totdebtlowinc if year==`year' & imp==`imp'
			mat tot`year'debtlow`imp' =(e(b))
			if `imp'==1 {
				local tot`year'low "tot`year'debtlow`imp'"
			}
			else {
				local tot`year'low "`tot`year'low'+tot`year'debtlow`imp'"
			}
		}
		disp "`tot`year'low'"
		qui mat tot`year'low=(`tot`year'low')/`im'
		qui mat colnames tot`year'low = totdebt
		qui mat rownames tot`year'low = `year'
	}
	*
	mat totlow=(tot1989low \ tot1992low \ tot1995low \ tot1998low \ tot2001low \ tot2004low \ tot2007low \ tot2010low \ tot2013low)
	mat list totlow
	gen totaldebtlow=.
	forvalues i=1/9 {
		replace totaldebtlow=totlow[`i',1]/10^9 in `i'
	}
	*checks
		sum totaldebt totaldebtlow
	*plot aggregate debt
```

```
capture drop lowshare
gen lowshare=totaldebtlow/totaldebt*100
replace lowshare=. in 1
replace lowshare=. in 2
twoway line lowshare yearplot, scheme(sj) ///
legend(subtitle("")) ytit("% of aggregate total debt", height(8)) xlabel(1989 1992 1995 1998 2001 2004 2007 2010 2013) ///
graphregion(fcolor(white))
graph export "$graphout\7 debt of hh with neg income shocks.tif", as(tif) replace
```

```
***********************************************************************
********Picture 8: Which part of the income distribution**************
***********************************************************************
*Define income percentiles
	local iter=0
	gen incperc=.
	forvalues year=1989(3)2013 {
		forvalues imp=1(1)5 {
			xtile inc`year'perc`imp'=income if year==`year' & imp==`imp' [pweight=wgt], n(100)
			replace incperc=inc`year'perc`imp' if year==`year' & imp==`imp'
			local iter=`iter'+1
			disp `iter'/45
		}
	}
*The Stata box plot command
	*the box displays the 25th and 75th percentile (which is the inter quartile range, IQR)
	*the median is in the middle
	*the whiskers extend 1.5 IQRs in each direction.
*produce box plot for white college
	graph box incper [fw=wgtround] if edcl<4 & race==1, over(year) ytit("income percentiles", height(8)) ///
	graphregion(fcolor(white)) title("non-college, white") scheme(sj)
	graph export "$graphout\8.1 nocoll white on inc dist.tif", as(tif) replace
*produce box plot for black college
	graph box incper [fw=wgtround] if edcl<4 & race==2, over(year) ytit("income percentiles", height(8)) ///
	graphregion(fcolor(white)) title("non-college, black") scheme(sj)
	graph export "$graphout\8.2 nocoll black on inc dist.tif", as(tif) replace
```

```
*produce box plot for white college
	graph box incper [fw=wgtround] if edcl==4 & race==1, over(year) ytit("income percentiles", height(8)) ///
	graphregion(fcolor(white)) title("college, white") scheme(sj)
	graph export "$graphout\8.3 col white on inc dist.tif", as(tif) replace
*produce box plot for black college
	graph box incper [fw=wgtround] if edcl==4 & race==2, over(year) ytit("income percentiles", height(8)) ///
	graphregion(fcolor(white)) title("college, black") scheme(sj)
	graph export "$graphout\8.4 col black on inc dist.tif", as(tif) replace
```

Appendix VII

This Appendix contains the code for the analysis presented in chapter 4.

```
**********************************************************************************************************
************************Section 1: Load data & define paths******************************************
**********************************************************************************************************

*load data
	clear all
	cd "C:\Users\Rafael\Documents\Data\SCF Data"
	use "estimation data v9.dta", replace

*output directories
	global outeleventhround "C:\Users\Rafael\Dropbox\Diss\SCF\results\eleventh round"

*only use years prior to the crisis
	drop if year>2007
	drop dum2010 dum2013

*set up complex survey design
	mi svyset [pw=wgt], bsrweight(repwgt1-repwgt999) vce(bootstrap)

*define inter race-edu group percentiles
	gen incperc_re=(1-headre)

*order dataset
	gen test=1-head
	order head test incperc, last


***************************************************************************************************************************
**********************Section 1A: Descriptive analysis of (d.D)/Y ************************************************
***************************************************************************************************************************

*count zero cases
forvalues i=1/5 {
```

```
	count if deltad2y==0 & _mi_m==`i'
}
*

*
count if debt==0 // 5679
count if debt==0 & debtlag2==0 // 5679

*check out availability
count if deltad2y==. & _mi_m!=0 //748
count if deltad2y==. & _mi_m==1 //151
count if deltad2y==. & _mi_m==2 //148
count if deltad2y==. & _mi_m==3 //152

*check out if rich households are more likely to borrowr or not
total wgt if deltad2y>0 & year==2004 & _mi_m==3
total wgt if deltad2y>0 & year==2004 & _mi_m==3 & incperc>=96 // 1.1 mio out of 16.8 = 6.55%
total wgt if deltad2y<=0 & year==2004 & _mi_m==3
total wgt if deltad2y<=0 & year==2004 & _mi_m==3 & incperc>=96 // 4.1 mio out of 92.4 = 4.44%
total wgt if deltad2y>0 & year==1995 & _mi_m==3
total wgt if deltad2y>0 & year==1995 & _mi_m==3 & incperc>=96 // 4.88%
total wgt if deltad2y<=0 & year==1995 & _mi_m==3
total wgt if deltad2y<=0 & year==1995 & _mi_m==3 & incperc>=96 // 4.9%
total wgt if deltad2y>0 & year==2001 & _mi_m==3
total wgt if deltad2y>0 & year==2001 & _mi_m==3 & incperc>=96 // 4.86%
total wgt if deltad2y<=0 & year==2001 & _mi_m==3
total wgt if deltad2y<=0 & year==2001 & _mi_m==3 & incperc>=96 // 4.96%
*seems only in 2004 there was a big difference

*plot histograms
sum deltad2y [fw=wgtround],d //so on average people take on 31% of their income in debt. Huge outliers are problem
count if deltad2y==0 & _mi_m!=0 //42.5k (~8.5k per implicate across all years)
sum deltad2y [fw=wgtround] if deltad2y>0,d //those who take on debt take on 20% at the median
sum deltad2y [fw=wgtround] if deltad2y<0,d //those who pay back pay 13% (median 8.8%)
count if deltad2y>10 & deltad2y!=. //89
count if deltad2y<-10 & deltad2y!=. //18
count if deltad2y<-5 & deltad2y!=. //43
count if deltad2y<-2.5 & deltad2y!=. //147
```

```
hist deltad2y if deltad2y<10 & deltad2y>-5 & deltad2y!=. [fw=wgtround]
sum incperc if (deltad2y>10 | deltad2<-2.5) & deltad2y!=.,d //these extreme outliers are from the very lower end of the income distribution
sum deltad2y if (deltad2y>10 | deltad2<-2.5) & deltad2y!=.,d //
*still quite skewed
count if deltad2y>5 & deltad2y!=. //270
count if deltad2y>2.5 & deltad2y!=. //900
count if deltad2y<-2.50 & deltad2y!=. //147
count if deltad2y<-1.50 & deltad2y!=. //242
hist deltad2y if deltad2y<2.50 & deltad2y>-1.50 & deltad2y!=. [fw=wgtround]
hist deltad2y if deltad2y<1 & deltad2y>-1 & deltad2y!=. [fw=wgtround]
*needs to be split
hist deltad2y if deltad2y<4 & deltad2y>0 & deltad2y!=. [fw=wgtround]
hist deltad2y if deltad2y<=0 & deltad2y>-2.50 & deltad2y!=. [fw=wgtround]
hist deltad2y if deltad2y<=0 & deltad2y>-2.50 & deltad2y!=. & debt!=0 [fw=wgtround]
hist deltad2y if deltad2y<0 & deltad2y>-2.50 & deltad2y!=. [fw=wgtround]
hist deltad2y if deltad2y<4 & deltad2y>-2.50 & deltad2y!=0 [fw=wgtround]
hist deltad2y if deltad2y<2 & deltad2y>-1.50 & deltad2y!=0 [fw=wgtround]
*So there is a massive amount of people with d.D=0

*who are the households with the outliers
*check income dist
        sum incper if deltad2y>2 & deltad2y!=.,d
        sum incper if deltad2y>4 & deltad2y!=.,d
        sum incper if deltad2y<-1 & deltad2y!=.,d //these are the poor median perc 19
        sum incper if deltad2y<-1.5 & deltad2y!=.,d //these are the poor median perc 10
        sum incper if deltad2y<-2.5 & deltad2y!=.,d //these are the poor median perc 4
        *So the observations with huge negative changes are the very poor!!
*check indebtedness
        sum debt if deltad2y>0,d
        sum debt if deltad2y>2 & deltad2y!=.,d
        sum debt if deltad2y>4 & deltad2y!=.,d
        sum debt if deltad2y<0,d
        sum debt if deltad2y<-1 & deltad2y!=.,d
        sum debt if deltad2y<-1.5 & deltad2y!=.,d
        sum debt if deltad2y<-2.5 & deltad2y!=.,d
        *in both cases the extreme values are the highly indebted households
        *especially with positive outliers their median debt is almost 4 times the
        *sample (only positive ones). With negative outliers not even twice
```

```
count if dnegall>4 | dpos>4   //20592
count if dnegall>4 //18041
count if dneg>4 //9462
count if dpos>4 //2551

*count aggregate d.D for each year
total ddebt2 [fw=wgtround] if year==1995 & _mi_m==1 //106 bn
total ddebt2 [fw=wgtround] if year==1998 & _mi_m==1 //113 bn
total ddebt2 [fw=wgtround] if year==2001 & _mi_m==1 //109 bn
total ddebt2 [fw=wgtround] if year==2004 & _mi_m==1 //697 bn
total ddebt2 [fw=wgtround] if year==2007 & _mi_m==1 //329 bn

*check out the asymetric nature of income and housing effects between
*the borrowing and non-borrowing sample
*full sample
	twoway scatter deltad2y ihs_income2 if year==1998 & _mi_m==3 & dperc3>4 & deltad2y>-2.5 & deltad2y<5, scheme(sj) ///
	|| lfit deltad2y ihs_income2 if year==1998 & _mi_m==3 & dperc3>4 & deltad2y>-2.5 & deltad2y<5
*borrowing
	twoway scatter deltad2y ihs_income2 if year==1998 & _mi_m==3 & dperc3>4 & deltad2y>0 & deltad2y<5, scheme(sj) ///
	|| lfit deltad2y ihs_income2 if year==1998 & _mi_m==3 & dperc3>4 & deltad2y>0 & deltad2y<5
*non-borrowing
	twoway scatter deltad2y ihs_income2 if year==1998 & _mi_m==3 & dperc3>4 & deltad2y>-1 & deltad2y<=0, scheme(sj) ///
	|| lfit deltad2y ihs_income2 if year==1998 & _mi_m==3 & dperc3>4 & deltad2y>-1 & deltad2y<=0



***********************************************************************************************
*****************************Section 2: sample split (Table 1)******************************
***********************************************************************************************


**************Table 1: with top1re, full, neg, and positive

*1) only positive, no interaction
	*run regression
		mi estimate, saving(rratio) post esampvaryok vceok: svy: reg deltad2y ihs_income2 ihs_avinctop1re ///
		ihs_HW dumHW ihs_REP dREP ihs_fin dfin ihs_debtlag2 kidsd age age2 eduaux##raceaux2 ///
```

```
		lf lowinc i.married turndown  ///
		dum1995 dum2001 dum2004 dum2007 if deltad2y>0 & deltad2y<5

		mi predict ratio_hat using rratio
		mi xeq 1: plot ratio_hat


		mi estimate, saving(level) post esampvaryok vceok: svy: reg ddebt2 ihs_income2 ihs_avinctop1re ///
		ihs_HW dumHW ihs_REP dREP ihs_fin dfin ihs_debtlag2 kidsd age age2 eduaux##raceaux2 ///
		lf lowinc i.married turndown  ///
		dum1995 dum2001 dum2004 dum2007 if deltad2y>0 & deltad2y<5

		///does not work cause predict only calculates it for 0 implicate!
		mi predict level_hat using level
		gen res_level=ddebt2-level_hat
		mi xeq 1: scatter level_hat ddebt2

		scatter ddebt2 res_level


	*save results
		estimates store tab1spec1
	*produce average R^2
		scalar r2 = 0
		qui mi xeq 1/5: reg deltad2y ihs_income2 ihs_avinctop1re ///
		ihs_HW dumHW ihs_REP dREP ihs_fin dfin ihs_debtlag2 kidsd age age2 eduaux##raceaux2 ///
		lf lowinc i.married turndown  ///
		dum1995 dum2001 dum2004 dum2007 [pw=wgt] if deltad2y>0 & deltad2y<5; scalar r2=r2+e(r2)
		scalar r2=r2/5
		sca list r2
*2) negative and zero d.D
	*run regression
		mi estimate, post esampvaryok vceok: svy: reg deltad2y ihs_income2 ihs_avinctop1re ///
		ihs_HW dumHW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 eduaux##raceaux2 ///
		lf lowinc i.married turndown  ///
		dum1995 dum2001 dum2004 dum2007 if deltad2y>-2.5 & deltad2y<=0
	*save results
		estimates store tab1spec2
```

```
		*produce average R^2
			scalar r2 = 0
			qui mi xeq 1/5: reg deltad2y ihs_income2 ihs_avinctop1re ///
			ihs_HW dumHW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 eduaux##raceaux2 ///
			lf lowinc i.married turndown  ///
			dum1995 dum2001 dum2004 dum2007 [pw=wgt] if deltad2y>-2.5 & deltad2y<=0; scalar r2=r2+e(r2)
			scalar r2=r2/5
			sca list r2
*3) full sample
		*run regression
			mi estimate, post esampvaryok vceok: svy: reg deltad2y ihs_income2 ihs_avinctop1re ///
			ihs_HW dumHW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 eduaux##raceaux2 ///
			lf lowinc i.married turndown  ///
			dum1995 dum2001 dum2004 dum2007 if deltad2y>-2.5 & deltad2y<5
		*save results
			estimates store tab1spec3
		*produce average R^2 (don't use bootstrap but weights instead -> quicker)
			scalar r2 = 0
			qui mi xeq 1/5: reg deltad2y ihs_income2 ihs_avinctop1re ///
			ihs_HW dumHW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 eduaux##raceaux2 ///
			lf lowinc i.married turndown  ///
			dum1995 dum2001 dum2004 dum2007 [pw=wgt] if deltad2y>-2.5 & deltad2y<5; scalar r2 = r2 + e(r2)
			scalar r2=r2/5
			sca list r2
*4) only positive, Y~ and HW interacted
		*run regression
			mi estimate, post esampvaryok vceok: svy: reg deltad2y ihs_income2 i.dumHW#c.ihs_avinctop1re ///
			ihs_HW dumHW ihs_REP dREP ihs_fin dfin ihs_debtlag2 kidsd age age2 dumHW##eduaux##raceaux2 ///
			lf lowinc i.married turndown   ///
			dum1995 dum2001 dum2004 dum2007 if deltad2y>0 & deltad2y<5
		*save results
			estimates store tab1spec4
		*produce average R^2
			scalar r2 = 0
			qui mi xeq 1/5: reg deltad2y ihs_income2 i.dumHW#c.ihs_avinctop1re ///
			ihs_HW dumHW ihs_REP dREP ihs_fin dfin ihs_debtlag2 kidsd age age2 dumHW##eduaux##raceaux2 ///
			lf lowinc i.married turndown   ///
			dum1995 dum2001 dum2004 dum2007 [pw=wgt] if deltad2y>0 & deltad2y<5; scalar r2=r2+e(r2)
```

```
		scalar r2=r2/5
		sca list r2
*5) negative, Y~ and HW interacted
	*run regression
		mi estimate, post esampvaryok vceok: svy: reg deltad2y ihs_income2 i.dumHW#c.ihs_avinctop1re ///
		ihs_HW dumHW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 dumHW##eduaux##raceaux2 ///
		lf lowinc i.married turndown   ///
		dum1995 dum2001 dum2004 dum2007 if deltad2y>-2.5 & deltad2y<=0
	*save results
		estimates store tab1spec5
	*produce average R^2
		scalar r2 = 0
		qui mi xeq 1/5: reg deltad2y ihs_income2 i.dumHW#c.ihs_avinctop1re ///
		ihs_HW dumHW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 dumHW##eduaux##raceaux2 ///
		lf lowinc i.married turndown   ///
		dum1995 dum2001 dum2004 dum2007 [pw=wgt] if deltad2y>-2.5 & deltad2y<=0; scalar r2=r2+e(r2)
		scalar r2=r2/5
		sca list r2
*export to Excel
	estout tab1spec* using "$outeleventhround\tab1.csv", ///
	cells( ///
		b(fmt(5)) ///
		se(star par('( )) fmt(3)) ) ///
	starlevels( * 0.1 ** 0.05 *** 0.01)  ///
	legend label ///
	varlabels( ///
		ihs_income2 ihs(Y) ///
		ihs_avinctop5re ihs(Y~) ///
		ihs_HW ihs(HW) ///
		ihs_REP ihs(REP) ///
		ihs_fin ihs(FW) ///
		ihs_debtlag2 ihs(Dt-1) ///
		kidsd kids ///
		age AGE ///
		age2 AGE^2) ///
	stats( ///
		N_mi F_mi p_mi rvi_avg_mi fmi_max_mi df_avg_mi, fmt(0 0 2 2 2 0) ///
		label(N F-stat Fpval RVI FMI avDF)) replace
```

```
*run specification without PP for answer to Bezemer (based on 4)
	*run regression
		mi estimate, post esampvaryok vceok: svy: reg deltad2y ihs_income2 ihs_avinctop1re ///
		ihs_fin dfin ihs_debtlag2 kidsd age age2 eduaux##raceaux2 ///
		lf lowinc i.married turndown   ///
		dum1995 dum2001 dum2004 dum2007 if deltad2y>0 & deltad2y<5
	*save results
		estimates store tab1spec4b
	*produce average R^2
		scalar r2 = 0
		qui mi xeq 1/5: reg deltad2y ihs_income2 ihs_avinctop1re ///
		ihs_fin dfin ihs_debtlag2 kidsd age age2 eduaux##raceaux2 ///
		lf lowinc i.married turndown   ///
		dum1995 dum2001 dum2004 dum2007 [pw=wgt] if deltad2y>0 & deltad2y<5; scalar r2=r2+e(r2)
		scalar r2=r2/5
		sca list r2
*include X301 (do you expect the US economy within the next 5 years to perform
*better (1), worse (2) or about the same (3)
*Because it is about a measure of optimising create dummy with X301==1
	*again regression based on 4 from above
	mi passive: gen optimism=1 if X301==1
	mi passive: replace optimism=0 if X301==2 | X301==3
	mi estimate, post esampvaryok vceok: svy: reg deltad2y ihs_income2 i.dumHW#c.ihs_avinctop1re ///
	ihs_HW dumHW ihs_REP dREP ihs_fin dfin ihs_debtlag2 kidsd age age2 dumHW##eduaux##raceaux2 ///
	lf lowinc i.married turndown optimism  ///
	dum1995 dum2001 dum2004 dum2007 if deltad2y>0 & deltad2y<5
*include interacted optimism dummy (based on 4)
*idea is to separate collateral arguments (HW) from Minsky arguments (HW*optimism)
	*run regression
		mi estimate, post esampvaryok vceok: svy: reg deltad2y ihs_income2 i.dumHW#c.ihs_avinctop1re ///
		i.optimism#c.ihs_HW dumHW ihs_REP dREP ihs_fin dfin ihs_debtlag2 kidsd age age2 dumHW##eduaux##raceaux2 ///
		lf lowinc i.married turndown optimism  ///
		dum1995 dum2001 dum2004 dum2007 if deltad2y>0 & deltad2y<5
	*HW is not different between optimistic and non-optimistic households
	*the dummy itself is not statistically significant either
*run spec 5 without zeros
```

```
			mi estimate, post esampvaryok vceok: svy: reg deltad2y ihs_income2 i.dumHW#c.ihs_avinctop1re ///
			ihs_HW dumHW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 dumHW##eduaux##raceaux2 ///
			lf lowinc i.married turndown   ///
			dum1995 dum2001 dum2004 dum2007 if deltad2y>-2.5 & deltad2y<0


*as check (Table 1B): report result when splitted between owners and non-owners
*BTW: there are 2314 observations which report houses=0 and have other real estate
*1) only positive, owners only
		*run regression
			mi estimate, post esampvaryok vceok: svy: reg deltad2y ihs_income2 ihs_avinctop1re ///
			ihs_HW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 eduaux##raceaux2 ///
			lf lowinc i.married turndown   ///
			dum1995 dum2001 dum2004 dum2007 if deltad2y>0 & deltad2y<5 & dumHW==1
		*save results
			estimates store tab1bspec1
*2) only positive, non-owners only
		*run regression
			mi estimate, post esampvaryok vceok: svy: reg deltad2y ihs_income2 ihs_avinctop1re ///
			ihs_HW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 eduaux##raceaux2 ///
			lf lowinc i.married turndown   ///
			dum1995 dum2001 dum2004 dum2007 if deltad2y>0 & deltad2y<5 & dumHW==0
		*save results
			estimates store tab1bspec2
*3) negative and zero d.D, owners
		*run regression
			mi estimate, post esampvaryok vceok: svy: reg deltad2y ihs_income2 ihs_avinctop1re ///
			ihs_HW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 eduaux##raceaux2 ///
			lf lowinc i.married turndown   ///
			dum1995 dum2001 dum2004 dum2007 if dperc3>4 & deltad2y>-2.5 & deltad2y<=0 & dumHW==1
		*save results
			estimates store tab1bspec3
*4) negative and zero d.D, non-owners
		*run regression
			mi estimate, post esampvaryok vceok: svy: reg deltad2y ihs_income2 ihs_avinctop1re ///
			ihs_HW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 eduaux##raceaux2 ///
			lf lowinc i.married turndown   ///
			dum1995 dum2001 dum2004 dum2007 if dperc3>4 & deltad2y>-2.5 & deltad2y<=0 & dumHW==0
```

```
	*save results
		estimates store tab1bspec4
*export to Excel
	estout tab1bspec* using "$outeleventhround\tab1_b.csv", ///
	cells( ///
		b(fmt(5)) ///
		se(star par('( )) fmt(3)) ) ///
	starlevels( * 0.1 ** 0.05 *** 0.01)  ///
	legend label ///
	varlabels( ///
		ihs_income2 ihs(Y) ///
		ihs_avinctop5re ihs(Y~) ///
		ihs_HW ihs(HW) ///
		ihs_REP ihs(REP) ///
		ihs_fin ihs(FW) ///
		ihs_debtlag2 ihs(Dt-1) ///
		kidsd kids ///
		age AGE ///
		age2 AGE^2) ///
	stats( ///
		N_mi F_mi p_mi rvi_avg_mi fmi_max_mi df_avg_mi, fmt(0 0 2 2 2 0) ///
		label(N F-stat Fpval RVI FMI avDF)) replace



****************************************************************************************
**********************Section 3: compare different Y~ measures (Table 2)************************
****************************************************************************************


********************Table 2a: more Y~ measures
	local iter=0
	foreach var in ihs_incp99re ihs_incp95re ihs_incp90re ihs_incp80re ihs_avinctop1re ihs_avinctop5re ihs_avinctop10re ihs_avinctop20re headre {

		local iter=`iter'+1
		noi disp "`var'"
		foreach group in pos neg {
			if "`group'"=="pos" {
```

```
                local lower=0
                local upper=5
            }
            else if "`group'"=="neg" {
                local lower=-2.5
                local upper "=0"
            }
            mi estimate, post esampvaryok vceok: svy: reg deltad2y ihs_income2 i.dumHW#c.`var' ///
            ihs_HW dumHW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 dumHW##eduaux##raceaux2 ///
            lf lowinc i.married turndown  ///
            dum1995 dum2001 dum2004 dum2007 if deltad2y>`lower' & deltad2y<`upper'
            estimates store tab2a`group'`iter'
        }
    }
*export to Excel
    foreach group in pos neg {
        estout tab2a`group'* using "$outeleventhround\tab2a`group'.csv", ///
        cells( ///
            b(fmt(5)) ///
            se(star par('( )) fmt(3)) ) ///
        starlevels( * 0.1 ** 0.05 *** 0.01)  ///
        legend label ///
        varlabels( ///
            ihs_income2 ihs(Y) ///
            ihs_avinctop5re ihs(Y~) ///
            ihs_HW ihs(HW) ///
            ihs_REP ihs(REP) ///
            ihs_fin ihs(FW) ///
            ihs_debtlag2 ihs(Dt-1) ///
            kidsd kids ///
            age AGE ///
            age2 AGE^2) ///
        stats( ///
            N_mi F_mi p_mi rvi_avg_mi fmi_max_mi df_avg_mi, fmt(0 0 2 2 2 0) ///
            label(N F-stat Fpval RVI FMI avDF)) replace
    }
    *
```

```
*********************Table 2b: regio measures on positive sample
	local iter=0
	foreach var in ihs_incp99re ihs_incp99regio ihs_incp95re ihs_incp95regio ihs_avinctop1re ihs_avinctop1regio ///
	ihs_avinctop5re ihs_avinctop5regio headre headregio {

		local iter=`iter'+1
		noi disp "`var'"
		foreach group in pos neg {
			if "`group'"=="pos" {
				local lower=0
				local upper=5
			}
			else if "`group'"=="neg" {
				local lower=-2.5
				local upper "=0"
			}
			mi estimate, post esampvaryok vceok: svy: reg deltad2y ihs_income2 i.dumHW#c.`var' ///
			ihs_HW dumHW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 dumHW##eduaux##raceaux2 ///
			lf lowinc i.married turndown i.X30074 ///
			dum1995 if deltad2y>`lower' & deltad2y<`upper' & year<=1998
			estimates store tab2b`group'`iter'
		}
	}
*export to Excel
	foreach group in pos neg {
		estout tab2b`group'* using "$outeleventhround\tab2b`group'.csv", ///
		cells( ///
			b(fmt(5)) ///
			se(star par('( )) fmt(3)) ) ///
		starlevels( * 0.1 ** 0.05 *** 0.01)  ///
		legend label ///
		varlabels( ///
			ihs_income2 ihs(Y) ///
			ihs_avinctop5re ihs(Y~) ///
			ihs_HW ihs(HW) ///
			ihs_REP ihs(REP) ///
			ihs_fin ihs(FW) ///
```

```
                ihs_debtlag2 ihs(Dt-1) ///
                kidsd kids ///
                age AGE ///
                age2 AGE^2) ///
        stats( ///
                N_mi F_mi p_mi rvi_avg_mi fmi_max_mi df_avg_mi, fmt(0 0 2 2 2 0) ///
                label(N F-stat Fpval RVI FMI avDF)) replace
}
*

*spec with edu-race not including regional dummies
mi estimate, post esampvaryok vceok: svy: reg deltad2y ihs_income2 i.dumHW#c.ihs_avinctop1re ///
ihs_HW dumHW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 dumHW##eduaux##raceaux2 ///
lf lowinc i.married turndown ///
dum1995 if deltad2y>0 & deltad2y<5 & year<=1998

*spec with region dummies, not including all edu-race interactions
mi estimate, post esampvaryok vceok: svy: reg deltad2y ihs_income2 i.dumHW#c.ihs_avinctop1regio ///
ihs_HW dumHW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 i.eduaux i.raceaux2 ///
lf lowinc i.married turndown i.X30074 ///
dum1995 if deltad2y>0 & deltad2y<5 & year<=1998


mi estimate, post esampvaryok vceok: svy: reg deltad2y ihs_income2 i.dumHW#c.headre ///
ihs_HW dumHW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 dumHW##eduaux##raceaux2 ///
lf lowinc i.married turndown ///
dum1995 if deltad2y>0 & deltad2y<5 & year<=1998 & ihs_avinctop1re!=.



*******************************************************************************************
**************************Section 4: effect size computation (Table 3)****************************
********************************************************************************************


*************************************************************
```

```
*********all baseline effects for avinctop1re****************
*************************************************************


*run baseline regressions (without replicate weights) to extract coefficient matrices
*NOTE: If "all" is also run then the neg and pos yhats are overwritten. Thus this
	*	section needs to be run either with pos and neg or ONLY with all without the
	*	other two!
foreach group in pos neg /*all*/ {
	noi disp " `group'"
	if "`group'"=="pos" {
		local lower=0
		local upper=5
	}
	else if "`group'"=="neg" {
		local lower=-2.5
		local upper "=0"
	}
	else if "`group'"=="all" {
		local lower=-2.5
		local upper=5
	}
	*run regression
		mi estimate, post esampvaryok vceok: reg deltad2y ihs_income2 i.dumHW#c.ihs_avinctop1re /// NOTE: adjust top 1% vs 5%
		ihs_HW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 dumHW##eduaux##raceaux2 ///
		lf lowinc i.married turndown b1998.year ///
		[pw=wgt] if deltad2y>`lower' & deltad2y<`upper'
	*save coefficient matrix
		mat coeff`group'=e(b)
		mat list coeff`group'
	*run again for each implicate
		mat coeffmat`group'=coeff`group'
		mi xeq 1/5: reg deltad2y ihs_income2 i.dumHW#c.ihs_avinctop1re /// NOTE: adjust top 1% vs 5%
		ihs_HW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 dumHW##eduaux##raceaux2 ///
		lf lowinc i.married turndown b1998.year ///
		[pw=wgt] if deltad2y>`lower' & deltad2y<`upper'; mat coeffmat`group'=(coeffmat`group' \ e(b))
		mat rownames coeffmat`group'=average imp1 imp2 imp3 imp4 imp5
		noi mat list coeffmat`group'
```

```
}
*compute aggregate income matrix but only for my actual sample (thus only white and black and -2.5<dhy<5)
	total income2 [fw=wgtround] if year==2004 & _mi_m==2 & (race==1 | race==2) & deltad2y>-2.5 & deltad2y<5
	total income2 [fw=wgtround] if year==2004 & _mi_m==2
	mat totinc=J(5,5,1)
	local iter=0
	qui forvalues year=1995(3)2007{
		local iter=`iter'+1
		mat aux=J(1,5,1)
		forvalues i=1(1)5 {
			total income2 [fw=wgtround] if year==`year' & _mi_m==`i' & (race==1 | race==2) & deltad2y>-2.5 & deltad2y<5
			mat totinc[`iter',`i']=e(b)
		}
	}
	mat colnames totinc=imp1 imp2 imp3 imp4 imp5
	mat rownames totinc=1995 1998 2001 2004 2007
	mat list totinc
********************************************************************************
*compute predicted values for each observation as % of total aggregate income
	*NOTE: If "all" is also run then the neg and pos yhats are overwritten. Thus this
	*	section needs to be run either with pos and neg or ONLY with all without the
	*	other two!
	*generate predicted values
		foreach var in inc HW0top1Y HW1top1Y hw dhw REP dREP fin dfin debtlag ddebtlag kidsd age age2 HW0college HW1college ///
		HW0black HW1black ///
		HW0collegeblack HW1collegeblack ///
		working lowinc notmarried turndown d1995 d2001 d2004 d2007 cons {
			foreach group in pos neg /*all*/ {
				foreach concept in rel abs {
					capture drop yh`group'_`var'`concept'
					gen yh`group'_`var'`concept'=.
				}
			}
		}
		qui foreach group in pos neg /*all*/ {
			noi disp " `group'"
			foreach concept in rel abs {
				if "`concept'"=="rel" {
```

```
		local inc " "
	}
	else if "`concept'"=="abs" {
		local inc "*income2"
	}
	if "`group'"=="pos" {
		local lower=0
		local upper=5
	}
	else if "`group'"=="neg" {
		local lower=-2.5
		local upper "=0"
	}
	else if "`group'"=="all" {
		local lower=-2.5
		local upper=5
	}
	forvalues i=1(1)5 {
		replace yh`group'_inc`concept'=			coeffmat`group'[1+`i',1]*ihs_income2`inc'		if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
		replace yh`group'_HW0top1Y`concept'=		coeffmat`group'[1+`i',2]*ihs_avinctop1re`inc'*(dumHW==0)		if _mi_m==`i' & deltad2y>`lower' &
deltad2y<`upper'
		replace yh`group'_HW1top1Y`concept'=		coeffmat`group'[1+`i',3]*ihs_avinctop1re`inc'*(dumHW==1)		if _mi_m==`i' & deltad2y>`lower' &
deltad2y<`upper'
		replace yh`group'_hw`concept'=			coeffmat`group'[1+`i',4]*ihs_HW`inc'			if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
		replace yh`group'_dhw`concept'=			coeffmat`group'[1+`i',15]*dumHW`inc'				if _mi_m==`i' & deltad2y>`lower' &
deltad2y<`upper'
		replace yh`group'_REP`concept'=			coeffmat`group'[1+`i',5]*ihs_REP`inc'			if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
		replace yh`group'_dREP`concept'=	coeffmat`group'[1+`i',6]*dREP`inc'				if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
		replace yh`group'_fin`concept'=			coeffmat`group'[1+`i',7]*ihs_fin`inc'			if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
		replace yh`group'_dfin`concept'=	coeffmat`group'[1+`i',8]*dfin`inc'				if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
		replace yh`group'_debtlag`concept'=	coeffmat`group'[1+`i',9]*ihs_debtlag2`inc'		if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
		replace yh`group'_ddebtlag`concept'=coeffmat`group'[1+`i',10]*ddebt2lag`inc'		if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'

		replace yh`group'_kidsd`concept'=	coeffmat`group'[1+`i',11]*kidsd`inc'			if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
		replace yh`group'_age`concept'=			coeffmat`group'[1+`i',12]*age`inc'				if _mi_m==`i' & deltad2y>`lower' &
deltad2y<`upper'
		replace yh`group'_age2`concept'=	coeffmat`group'[1+`i',13]*age2`inc'				if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
```

```
replace yh`group'_HW0college`concept'=     coeffmat`group'[1+`i',17]*eduaux`inc'*(dumHW==0)     if _mi_m==`i' & deltad2y>`lower' &
deltad2y<`upper'
replace yh`group'_HW1college`concept'=     coeffmat`group'[1+`i',21]*eduaux`inc'*(dumHW==1)     if _mi_m==`i' & deltad2y>`lower' &
deltad2y<`upper'
replace yh`group'_HW0black`concept'=     coeffmat`group'[1+`i',23]*(race==2)`inc'*(dumHW==0)     if _mi_m==`i' & deltad2y>`lower' &
deltad2y<`upper'
replace yh`group'_HW1black`concept'=     coeffmat`group'[1+`i',27]*(race==2)`inc'*(dumHW==1)     if _mi_m==`i' & deltad2y>`lower' &
deltad2y<`upper'
replace yh`group'_HW0collegeblack`concept'=coeffmat`group'[1+`i',31]*(race==2)*(eduaux==1)`inc'*(dumHW==0) if _mi_m==`i' & deltad2y>`lower' &
deltad2y<`upper'
replace yh`group'_HW1collegeblack`concept'=coeffmat`group'[1+`i',39]*(race==2)*(eduaux==1)`inc'*(dumHW==1) if _mi_m==`i' & deltad2y>`lower' &
deltad2y<`upper'
replace yh`group'_working`concept'=  coeffmat`group'[1+`i',40]*lf`inc'     if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
replace yh`group'_lowinc`concept'=   coeffmat`group'[1+`i',41]*lowinc`inc'     if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
replace yh`group'_notmarried`concept'=coeffmat`group'[1+`i',43]*(married==2)`inc' if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
replace yh`group'_turndown`concept'=coeffmat`group'[1+`i',44]*turndown`inc'     if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
replace yh`group'_d1995`concept'=    coeffmat`group'[1+`i',45]*(year==1995)`inc' if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
replace yh`group'_d2001`concept'=    coeffmat`group'[1+`i',47]*(year==2001)`inc'     if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
replace yh`group'_d2004`concept'=    coeffmat`group'[1+`i',48]*(year==2004)`inc' if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
replace yh`group'_d2007`concept'=    coeffmat`group'[1+`i',49]*(year==2007)`inc' if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
replace yh`group'_cons`concept'=     coeffmat`group'[1+`i',50]`inc'     if _mi_m==`i' & deltad2y>`lower' &
deltad2y<`upper'
			}
		}
	}
*now check whether the model produces negative predicted values in the borrowing sample
*and positive predicted values in the non-borrowingsample
	foreach group in pos neg /*all*/ {
		foreach concept in rel abs {
			capture drop yh`group'_total`concept'
			gen yh`group'_total`concept'=yh`group'_inc`concept'
			foreach var in HW0top1Y HW1top1Y hw dhw REP dREP fin dfin debtlag ddebtlag kidsd age age2 /// NOTE: adjust top 1% vs 5%
			HW0college HW1college HW0black HW1black ///
			HW0collegeblack HW1collegeblack ///
			working lowinc notmarried turndown d1995 d2001 d2004 d2007 cons {
				replace yh`group'_total`concept'=yh`group'_total`concept'+yh`group'_`var'`concept'
			}
		}
```

```
}
*check missing
		sum yhpos_total*
		sum yhneg_total*
		sum yhall_total*
		sum yhpos_inc*
		sum deltad2y if deltad2y>0 & deltad2y<5
*so missing values are due to the dropped households of Y~
		*NOTE: adjust pos/neg and the deltad2y ranges accordingly
		count if yhpos_totalrel==. & _mi_m!=0 & deltad2y>0 & deltad2y<5 // 1.2k
		foreach var in inc HW0top1Y HW1top1Y hw dhw REP dREP fin dfin debtlag ddebtlag kidsd age age2 ///
		HW0college HW1college HW0black HW1black ///
		HW0collegeblack HW1collegeblack ///
		working lowinc notmarried turndown d1995 d2001 d2004 d2007 cons {
			count if yhpos_`var'rel==. & _mi_m!=0 & deltad2y>0 & deltad2y<5
		}
*positive first
		sum yhpos_totalrel if _mi_m!=0 & deltad2y>0 & deltad2y<5 //12.5k
		count if yhpos_totalrel<0 & _mi_m!=0 & deltad2y>0 & deltad2y<5 //1.6k
		count if yhpos_totalrel==0 & _mi_m!=0 & deltad2y>0 & deltad2y<5 //0 obs
		count if yhpos_totalrel>0 & _mi_m!=0 & deltad2y>0 & deltad2y<5 & yhpos_totalrel!=. //11k
		*12.1% are nonsensical
*negative
		sum yhneg_totalrel if _mi_m!=0 & deltad2y>-2.5 & deltad2y<=0 // 79.7k
		count if yhneg_totalrel>0 & _mi_m!=0 & deltad2y>-2.5 & deltad2y<=0 & yhneg_totalrel!=. //16.5k
		count if yhneg_totalrel<0 & _mi_m!=0 & deltad2y==0 & yhneg_totalrel!=. //19.2k
		*44.8% are nonsensical
		sum deltad2y if yhneg_totalrel>0 & deltad2y>-2.5 & deltad2y<=0 & yhneg_totalrel!=.,d //99% of them have d.D=0
		sum deltad2y if yhneg_totalrel<=0 & deltad2y>-2.5 & deltad2y<=0 & yhneg_totalrel!=.,d
		count if deltad2y==0 & deltad2y>-2.5 & deltad2y<=0 & yhneg_totalrel!=.
		*so these are about half the observations with no change in debt
		*overall (pos+neg) 59.9% make sense and 40.1% are non-sensical
*for the all case
		sum yhall_totalrel if _mi_m!=0 & deltad2y>-2.5 & deltad2y<5 // 92.2k
		count if yhall_totalrel>0 & _mi_m!=0 & deltad2y>-2.5 & deltad2y<=0 & yhall_totalrel!=. // 21.9k which is 6k more than in split
		count if yhall_totalrel<0 & _mi_m!=0 & deltad2y==0 & yhall_totalrel!=. // 19.7k
		count if yhall_totalrel<0 & _mi_m!=0 & deltad2y>0 & deltad2y<5 & yhall_totalrel!=. // 7.5k which is 6k more than in split
		*53.2% nonsensical predictions
```

```
*46.8% of all predicted cases make sense
*check whether the predictions hold in the cross section
*in the cross section
	*dependent variable (holds perfectly)
		sum yhpos_totalrel [fw=wgtround] if _mi_m==3 & deltad2y>0 & deltad2y<5
		sum deltad2y [fw=wgtround] if _mi_m==3 & deltad2y>0 & deltad2y<5 & avinctop1re!=.
	*dep var scaled back to absolute debt changes
		*positive section
			total yhpos_totalabs [fw=wgtround] if _mi_m==3 & deltad2y>0 & deltad2y<5 //2.74 tr
			total yhall_totalabs [fw=wgtround] if _mi_m==3 & deltad2y>0 & deltad2y<5 //1.88 tr
			total ddebt2 [fw=wgtround] if _mi_m==3 & deltad2y>0 & deltad2y<5 & avinctop1re!=. //2.7 tr
			total ddebt2 [fw=wgtround] if _mi_m==3 & deltad2y>0 & deltad2y<5  //2.86 tr
		*negative section
			total yhneg_totalabs [fw=wgtround] if _mi_m==3 & deltad2y>-2.5 & deltad2y<=0 //-1.86 tr (pos+neg=880bn)
			total yhall_totalabs [fw=wgtround] if _mi_m==3 & deltad2y>-2.5 & deltad2y<=0 //-0.81 tr		(pos+neg=1070bn)
			total ddebt2 [fw=wgtround] if _mi_m==3 & deltad2y>-2.5 & deltad2y<=0 & avinctop1re!=. //-1.91 tr (pos+neg=790bn)
			total ddebt2 [fw=wgtround] if _mi_m==3 & deltad2y>-2.5 & deltad2y<=0  //-1.99 tr (pos+neg=850bn)
*for individual years
	*dependent variable (almost perfect)
		sum yhpos_totalrel [fw=wgtround] if _mi_m==3 & deltad2y>0 & deltad2y<5 & year==2004
		sum deltad2y [fw=wgtround] if _mi_m==3 & deltad2y>0 & deltad2y<5 & avinctop1re!=. & year==2004
	*dep var scaled back to absolute debt changes
		total yhpos_totalabs [fw=wgtround] if _mi_m==3 & deltad2y>0 & deltad2y<5 & year==2004 //986 bn
		total ddebt2 [fw=wgtround] if _mi_m==3 & deltad2y>0 & deltad2y<5 & avinctop1re!=. & year==2004 //817 bn
*sum effects and express in % of total aggregate income
		qui foreach group in pos neg /*all*/ {
			noi disp " `group'"
			mat eff`group'=J(5,30,1)
			mat eff`group'dif=J(4,30,1)
			if "`group'"=="pos" {
				local lower=0
				local upper=5
				*local logic "yhpos_totalrel>0 & yhpos_totalrel!=."
			}
			else if "`group'"=="neg" {
				local lower=-2.5
				local upper "=0"
				*local logic "yhneg_totalrel<=0"
```

```
}
else if "`group'"=="all" {
		local lower=-2.5
		local upper=5
		*local logic "((deltad2y>0 & yhall_totalrel>0 & yhall_totalrel!=.) | (deltad2y<=0 & yhall_totalrel<=0))"
}
local iter=0
forvalues year=1995(3)2007 {
		local iter=`iter'+1
		noi disp "`year'"
		mat aux`year'=J(5,30,1)
		mat aux2dif`year'=J(5,30,1)
		forvalues i=1(1)5 {
				local iter2=0
				foreach var in inc HW0top1Y HW1top1Y hw dhw REP dREP fin dfin debtlag ddebtlag kidsd age age2 ///
				HW0college HW1college HW0black HW1black ///
				HW0collegeblack HW1collegeblack ///
				working lowinc notmarried turndown d1995 d2001 d2004 d2007 cons total {
						local iter2=`iter2'+1
						total yh`group'_`var'abs [fw=wgtround] if year==`year' & _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper' & avinctop1re!=. // & `logic' /*NOTE:
adjust top 1% vs 5%*/

						mat aux`year'[`i',`iter2']=e(b)
						if `year'>=1998 {
								mat auxdif`year'=(aux`year'-aux1995)
								mat aux2dif`year'[`i',`iter2']=auxdif`year'[`i',`iter2']/totinc[`iter',`i']
						}
				}
		}
		mat aux2=(J(1,5,1)*aux`year')/5
		forvalues j=1(1)30 {
				mat eff`group'[`iter',`j']=aux2[1,`j']
		}
		if `year'>=1998 {
				mat aux3dif=(J(1,5,1)*aux2dif`year')/5
				forvalues j=1(1)30 {
						mat eff`group'dif[`iter'-1,`j']=aux3dif[1,`j']
				}
		}
```

```
				}
				mat rownames eff`group'=1995 1998 2001 2004 2007
				mat rownames eff`group'dif=1998 2001 2004 2007
				mat colnames eff`group'=inc HW0top1Y HW1top1Y hw dhw REP dREP fin dfin debtlag ddebtlag kidsd age age2 ///
							HW0college HW1college HW0black HW1black ///
							HW0collegeblack HW1collegeblack ///
							working lowinc notmarried turndown d1995 d2001 d2004 d2007 cons total
				mat colnames eff`group'dif=inc HW0top1Y HW1top1Y hw dhw REP dREP fin dfin debtlag ddebtlag kidsd age age2 ///
							HW0college HW1college HW0black HW1black ///
							HW0collegeblack HW1collegeblack ///
							working lowinc notmarried turndown d1995 d2001 d2004 d2007 cons total
			}
			mat list aux
			mat list effpos
			mat list effneg
			mat list effall
		*run some checks
			mat test=J(1,5,1)
			mat aux3=J(5,1,1)
			qui forvalues i=1(1)5{
				total yhneg_incabs [fw=wgtround] if year==2007 & deltad2y>-2.5 & deltad2y<=0 & _mi_m==`i' & avinctop1re!=.
				mat test[1,`i']=e(b)/totinc[5,`i']
			}
			mat aux4=test*aux3/5
			mat list aux4
			*check how many positive observatins for REP are there
			count if REP>0 & REP!=. & _mi_m!=0 & deltad2y>-2.5 & deltad2y<=0
			count if yhneg_REPabs>0 & yhneg_REPabs!=. & _mi_m!=0 & deltad2y>-2.5 & deltad2y<=0
*export everything to Ecel
		xml_tab effpos, save("$outeleventhround\effects_pos.xml") replace
		xml_tab effneg, save("$outeleventhround\effects_neg.xml") replace
		*xml_tab effall, save("$outeleventhround\effects_all.xml") replace
		xml_tab effposdif, save("$outeleventhround\effects_posdif.xml") replace
		xml_tab effnegdif, save("$outeleventhround\effects_negdif.xml") replace
```

```
********************************************************************************
************all baseline effects for all other Y~ specifications ***************
********************************************************************************

qui foreach def in /*incp99re incp95re incp90re incp80re avinctop1re*/ avinctop5re /*avinctop10re avinctop20re*/ {
	foreach group in pos neg {
		noi disp " `group'"
		if "`group'"=="pos" {
			local lower=0
			local upper=5
		}
		else if "`group'"=="neg" {
			local lower=-2.5
			local upper "=0"
		}
	********************************************************************************
	*run baseline regressions (without replicate weights) to extract coefficient matrices
	*run regression
		mi estimate, post esampvaryok vceok: reg deltad2y ihs_income2 i.dumHW#c.ihs_`def' ///
		ihs_HW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 dumHW##eduaux##raceaux2 ///
		lf lowinc i.married turndown b1998.year ///
		[pw=wgt] if deltad2y>`lower' & deltad2y<`upper'
	*save coefficient matrix
		mat coeff`group'`def'=e(b)
		mat list coeff`group'`def'
	*run again for each implicate
		mat coeffmat`group'`def'=coeff`group'`def'
		mi xeq 1/5: reg deltad2y ihs_income2 i.dumHW#c.ihs_`def' /// NOTE: adjust top 1% vs 5%
		ihs_HW ihs_REP dREP ihs_fin dfin ihs_debtlag2 ddebt2lag kidsd age age2 dumHW##eduaux##raceaux2 ///
		lf lowinc i.married turndown b1998.year ///
		[pw=wgt] if deltad2y>`lower' & deltad2y<`upper'; mat coeffmat`group'`def'=(coeffmat`group'`def' \ e(b))
		mat rownames coeffmat`group'`def'=average imp1 imp2 imp3 imp4 imp5
		noi mat list coeffmat`group'`def'
	********************************************************************************
	*compute predicted values for each observation as % of total aggregate income
	*generate predicted values
		foreach var in inc HW0`def' HW1`def' hw dhw REP dREP fin dfin debtlag ddebtlag kidsd age age2 HW0college HW1college ///
		HW0black HW1black ///
```

```
		HW0collegeblack HW1collegeblack ///
		working lowinc notmarried turndown d1995 d2001 d2004 d2007 cons {
			foreach concept in abs {
				capture drop yh`group'_`var'`concept'
				gen yh`group'_`var'`concept'=.
			}
		}
		foreach concept in abs {
			if "`concept'"=="rel" {
				local inc " "
			}
			else if "`concept'"=="abs" {
				local inc "*income2"
			}
			forvalues i=1(1)5 {
				replace yh`group'_inc`concept'=			coeffmat`group'`def'[1+`i',1]*ihs_income2`inc'	if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_HW0`def'`concept'= coeffmat`group'`def'[1+`i',2]*ihs_avinctop1re`inc'*(dumHW==0)	if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_HW1`def'`concept'= coeffmat`group'`def'[1+`i',3]*ihs_avinctop1re`inc'*(dumHW==1)	if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_hw`concept'=			coeffmat`group'`def'[1+`i',4]*ihs_HW`inc'		if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_dhw`concept'=			coeffmat`group'`def'[1+`i',15]*dumHW`inc'			if _mi_m==`i' & deltad2y>`lower' &
deltad2y<`upper'
				replace yh`group'_REP`concept'=			coeffmat`group'`def'[1+`i',5]*ihs_REP`inc'		if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_dREP`concept'=	coeffmat`group'`def'[1+`i',6]*dREP`inc'			if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_fin`concept'=			coeffmat`group'`def'[1+`i',7]*ihs_fin`inc'		if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_dfin`concept'=	coeffmat`group'`def'[1+`i',8]*dfin`inc'		if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_debtlag`concept'=	coeffmat`group'`def'[1+`i',9]*ihs_debtlag2`inc'	if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_ddebtlag`concept'=coeffmat`group'`def'[1+`i',10]*ddebt2lag`inc'	if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'

				replace yh`group'_kidsd`concept'=	coeffmat`group'`def'[1+`i',11]*kidsd`inc'		if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_age`concept'=			coeffmat`group'`def'[1+`i',12]*age`inc'		if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_age2`concept'=	coeffmat`group'`def'[1+`i',13]*age2`inc'			if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_HW0college`concept'=		coeffmat`group'`def'[1+`i',17]*eduaux`inc'*(dumHW==0)		if _mi_m==`i' & deltad2y>`lower' &
deltad2y<`upper'
				replace yh`group'_HW1college`concept'=		coeffmat`group'`def'[1+`i',21]*eduaux`inc'*(dumHW==1)		if _mi_m==`i' & deltad2y>`lower' &
deltad2y<`upper'
				replace yh`group'_HW0black`concept'=		coeffmat`group'`def'[1+`i',23]*(race==2)`inc'*(dumHW==0)	if _mi_m==`i' & deltad2y>`lower' &
deltad2y<`upper'
```

```
				replace yh`group'_HW1black`concept'=		coeffmat`group'`def'[1+`i',27]*(race==2)`inc'*(dumHW==1)		if _mi_m==`i' & deltad2y>`lower' &
deltad2y<`upper'
				replace yh`group'_HW0collegeblack`concept'=coeffmat`group'`def'[1+`i',31]*(race==2)*(eduaux==1)`inc'*(dumHW==0) if _mi_m==`i' & deltad2y>`lower' &
deltad2y<`upper'
				replace yh`group'_HW1collegeblack`concept'=coeffmat`group'`def'[1+`i',39]*(race==2)*(eduaux==1)`inc'*(dumHW==1) if _mi_m==`i' & deltad2y>`lower' &
deltad2y<`upper'
				replace yh`group'_working`concept'=	coeffmat`group'`def'[1+`i',40]*lf`inc'			if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_lowinc`concept'=	coeffmat`group'`def'[1+`i',41]*lowinc`inc'		if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_notmarried`concept'=coeffmat`group'`def'[1+`i',43]*(married==2)`inc' if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_turndown`concept'=coeffmat`group'`def'[1+`i',44]*turndown`inc'		if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_d1995`concept'=	coeffmat`group'`def'[1+`i',45]*(year==1995)`inc' if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_d2001`concept'=	coeffmat`group'`def'[1+`i',47]*(year==2001)`inc'  if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_d2004`concept'=	coeffmat`group'`def'[1+`i',48]*(year==2004)`inc' if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_d2007`concept'=	coeffmat`group'`def'[1+`i',49]*(year==2007)`inc' if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
				replace yh`group'_cons`concept'=	coeffmat`group'`def'[1+`i',50]`inc'				if _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper'
			}
		}
		*produce total over all variables
			foreach concept in abs {
				capture drop yh`group'_total`concept'
				gen yh`group'_total`concept'=yh`group'_inc`concept'
				foreach var in HW0`def' HW1`def' hw dhw REP dREP fin dfin debtlag ddebtlag kidsd age age2 /// NOTE: adjust top 1% vs 5%
				HW0college HW1college HW0black HW1black ///
				HW0collegeblack HW1collegeblack ///
				working lowinc notmarried turndown d1995 d2001 d2004 d2007 cons {
					replace yh`group'_total`concept'=yh`group'_total`concept'+yh`group'_`var'`concept'
				}
			}
	******************************************************************************
	*sum effects and express in % of total aggregate income
		mat eff`group'`def'=J(5,30,1)
		mat eff`group'`def'dif=J(4,30,1)
		local iter=0
		forvalues time=1995(3)2007 {
			noi disp " `time'"
			local iter=`iter'+1
			mat aux`time'=J(5,30,1)
			mat aux2dif`time'=J(5,30,1)
```

```
		forvalues i=1(1)5 {
			local iter2=0
			foreach var in inc HW0`def' HW1`def' hw dhw REP dREP fin dfin debtlag ddebtlag kidsd age age2 HW0college HW1college ///
			HW0black HW1black  ///
			HW0collegeblack HW1collegeblack  ///
			working lowinc notmarried turndown d1995 d2001 d2004 d2007 cons total {
				local iter2=`iter2'+1
				total yh`group'_`var'abs [fw=wgtround] if year==`time' & _mi_m==`i' & deltad2y>`lower' & deltad2y<`upper' & `def'!=. // & `logic'
				mat aux`time'[`i',`iter2']=e(b)
				if `time'>=1998 {
					mat auxdif`time'=(aux`time'-aux1995)
					mat aux2dif`time'[`i',`iter2']=auxdif`time'[`i',`iter2']/totinc[`iter',`i']
				}
			}
		}
		mat aux2=(J(1,5,1)*aux`time')/5
		forvalues j=1(1)30 {
			mat eff`group'`def'[`iter',`j']=aux2[1,`j']
		}
		if `time'>=1998 {
			mat aux3dif=(J(1,5,1)*aux2dif`time')/5
			forvalues j=1(1)30 {
				mat eff`group'`def'dif[`iter'-1,`j']=aux3dif[1,`j']
			}
		}
	}
	mat rownames eff`group'`def'=1995 1998 2001 2004 2007
	mat rownames eff`group'`def'dif= 1998 2001 2004 2007
	mat colnames eff`group'`def'=inc HW0`def' HW1`def' hw dhw REP dREP fin dfin debtlag ddebtlag kidsd age age2 HW0college HW1college ///
			HW0black HW1black  ///
			HW0collegeblack HW1collegeblack  ///
			working lowinc notmarried turndown d1995 d2001 d2004 d2007 cons total
	mat colnames eff`group'`def'dif=inc HW0`def' HW1`def' hw dhw REP dREP fin dfin debtlag ddebtlag kidsd age age2 HW0college HW1college ///
			HW0black HW1black ///
			HW0collegeblack HW1collegeblack  ///
			working lowinc notmarried turndown d1995 d2001 d2004 d2007 cons total
****************************************************************************
*export everything to Ecel
```

```
			xml_tab eff`group'`def', save("$outeleventhround\effects_`group'_`def'.xml") replace
			xml_tab eff`group'`def'dif, save("$outeleventhround\effects_`group'_`def'dif.xml") replace
		}
}
*run some checks
	*all coeff matrices are defined
		mat list coeffmatposincp99re
		mat list coeffmatposavinctop10re
	*compute predicted values
		mat test=J(1,5,1)
		gen testpred=.
		forvalues i=1/5 {
			replace testpred=coeffmatposincp99re[`i'+1,2]*ihs_incp99re*income2 if _mi_m==`i' & deltad2y>0 & deltad2y<5
		}
	*aggregate and average over implicates
		mat effect=J(1,5,1)
		qui forvalues i=1/5 {
			total testpred [fw=wgtround] if year==1995 & _mi_m==`i' & deltad2y>0 & deltad2y<5 & ihs_incp99re!=.
			mat effect[1,`i']=e(b)
		}
		mat aux=(effect*J(5,1,1))/5
		mat list aux
	*compute diff scaled by aggregate income (set previous section with effect to 1995)
		mat effectaux=J(1,5,1)
		qui forvalues i=1/5 {
			total testpred [fw=wgtround] if year==2004 & _mi_m==`i' & deltad2y>0 & deltad2y<5 & ihs_incp99re!=.
			mat effectaux[1,`i']=e(b)
		}
		mat effdif=effectaux-effect
		mat effrel=J(1,5,1)
		forvalues i=1/5 {
			mat effrel[1,`i']=effdif[1,`i']/totinc[4,`i']
		}
		mat aux=(effrel*J(5,1,1))/5
		mat list aux
*check out all coeff matrices for individual implicates
foreach var in incp80re /*incp99re /*incp95re incp90re*/ incp80re /*avinctop5re*/ avinctop10re avinctop20re*/ {
	mat list coeffmatpos`var'
```

}
*for p99 the positive sample has positive and negative coeffs for Y~ (2 out of 5) (negative not)
*same holds true for p95 (2 out of 5)
*for p90 1 out of 5 implicates has negative coeff for Y~
*for p80 none is negative

Appendix VIII

This Appendix contains the code used for chapter 5

```
*******************************************************************************************
*******************************************************************************************
*************************INET Project Paper 2: Debt and Consumption*****************************
*******************************************************************************************
*******************************************************************************************


******************************************************************************************
******************************Section 1: Data Input********************************************
*******************************************************************************************

**Use the standard dataset
clear
use "C:\Users\Rafael\Google Drive\INET project\data\stata work files\data long v06.dta", clear

egen cntry = group(country)

**Set panel
xtset cntry year
tab cntry

**Generate log of all relevant variables
foreach var in x m inv gni y c ///
top10a top5a top1a top05a top10swtid dhy dhyd dbus wsfc intlrmix ///
oecdy ex exnom ulc pop credreg finreform ///
spmix dh pp ehii ///
gini_net gini_market share1 yd netlend rgdpo {
gen `var'l=ln(`var')
}
*

*real change in debt as % of real GDP
gen ddhy=(d.dh)/y
```

```
*nominal change in debt as % of disposable income
gen aux=dh*pc
gen ddhyd2=d.aux/(yd*pc)

*real change in debt as % of real disposable income
gen ddhyd=(d.dh)/yd

*time trend and country dummies
by cntry: gen time=_n
*tab country, generate(cnt)

*construct country dummies


*saving
gen savy=cl-yl
gen savyd=cl-ydl


*******************************************************************************************
********************************Section 2: Create Sample***************************************
*******************************************************************************************
**create a balanced panel
drop if year>2013

*OECD members: au at be ca chile cz dk ee fi fr de gr hu is ie israel it jp kr lu mx nl nz no pl pt sk si es se ch tr gb us
*not OECD members but in my sample: bg cy hr lt lv mk mt ro
*in OECD but not in my sample: chile, israel
*only use OECD members
drop if country == "bg" | country=="cy" | country == "hr" | country=="lt" | country=="lv" ///
| country == "mk" | country == "mt" | country == "ro"
*keep in mind that norway is a huge outlier in the share1 series

*check that the main variables are available for all countries of the panel
drop if c==. | yd==. | dh==. | pp==. | spmix==.

*This results in the familiar 18 country panel data set:
```

```
*at, au , be, ca, ch, de, dk, es, fi, fr, gb, ie, it, jp, nl, no, se, us

drop if year<1980 //only deletes 22 obserations

*generate time dummies
tab year, gen(dum)

tab year
tab country

*the numbers represent the countrd ID number:
*AT 1, CH 6, DK 10,  ES 12, IE 19, NL 30
*these countries have the following numbers of observations:
*AT: 14, CH: 13, DK: 19, ES: 18, IE: 12, NL: 24


*generate consumption and disposable income in PPPs
gen oldratio=old/pop
gen oldratiol=ln(oldratio)

*set path
global out "C:\Users\Rafael Wildauer\Dropbox\Diss\paper 2\results"


******************************************************************************************
***************************Section 3: Unit Root Tests*******************************************
******************************************************************************************

*relevant variables: ydPPP cPPP intlrmix oldratio dhPPP share1 gini_market ehii pp spmix credreg finreform
*baseline debt spec: xtivreg2 dhPPPl ydPPPl share1 ppl spmixl intlrmix oldratio finreform if country!="no" & country!="ie" ///
*                                       & country!="es" & country!="at" & country!="dk" & country!="ch" & year<2008
*baseline cons spec: xtivreg2 cPPPl ydPPPl share1 d.ppl d.spmixl intlrmix, fd cluster(country) noc


xtunitroot fisher d.dhPPPl if year<2008, dfuller lags(3) trend


*run ADF tests (levels)
```

```
foreach var in dhPPPl cPPPl ydPPPl ppl spmixl share1 gini_market ehii oldratio credreg finreform {
		xtunitroot fisher `var' if country!="no" & country!="ie" ///
		& country!="es" & country!="at" & country!="dk" & country!="ch" & year<2008, dfuller lags(3) trend
		mata: p1`var'=(st_numscalar("r(p_P)"),st_numscalar("r(p_L)"),st_numscalar("r(p_Z)"))
}
foreach var in intlrmix oldratio share1 gini_market ehii credreg finreform {
		xtunitroot fisher `var' if country!="no" & country!="ie" ///
		& country!="es" & country!="at" & country!="dk" & country!="ch" & year<2008, dfuller lags(3)
		mata: p2`var'=(st_numscalar("r(p_P)"),st_numscalar("r(p_L)"),st_numscalar("r(p_Z)"))
}
mata
all=(p1dhPPPl\p1cPPPl\p1ydPPPl\p1ppl\p1spmixl\p1share1\p1gini_market\p1ehii\p1oldratio\p1credreg\p1finreform ///
\p2intlrmix\p2oldratio\p2share1\p2gini_market\p2ehii\p2credreg\p2finreform)
st_matrix("all",all)
end
cd "C:\Users\Rafael Wildauer\Dropbox\Diss\paper 2\results"
xml_tab all, save("unit root tests level.xml") replace


*run ADF tests (levels, including the crisis)
foreach var in dhPPPl cPPPl ydPPPl ppl spmixl share1 gini_market ehii oldratio credreg finreform {
		xtunitroot fisher `var' if country!="no" & country!="ie" ///
		& country!="es" & country!="at" & country!="dk" & country!="ch", dfuller lags(3) trend
		mata: p1`var'=(st_numscalar("r(p_P)"),st_numscalar("r(p_L)"),st_numscalar("r(p_Z)"))
}
foreach var in intlrmix oldratio share1 gini_market ehii credreg finreform {
		xtunitroot fisher `var' if country!="no" & country!="ie" ///
		& country!="es" & country!="at" & country!="dk" & country!="ch", dfuller lags(3)
		mata: p2`var'=(st_numscalar("r(p_P)"),st_numscalar("r(p_L)"),st_numscalar("r(p_Z)"))
}
mata
all=(p1dhPPPl\p1cPPPl\p1ydPPPl\p1ppl\p1spmixl\p1share1\p1gini_market\p1ehii\p1oldratio\p1credreg\p1finreform ///
\p2intlrmix\p2oldratio\p2share1\p2gini_market\p2ehii\p2credreg\p2finreform)
st_matrix("all",all)
end
cd "C:\Users\Rafael Wildauer\Dropbox\Diss\paper 2\results"
xml_tab all, save("unit root tests level including crisis.xml") replace
```

```
*run ADF tests (diffs)
foreach var in dhPPPl cPPPl ydPPPl ppl spmixl share1 gini_market ehii oldratio credreg finreform {
	xtunitroot fisher d.`var' if country!="no" & country!="ie" ///
	& country!="es" & country!="at" & country!="dk" & country!="ch" & year<2008, dfuller lags(3) trend
	mata: p1`var'=(st_numscalar("r(p_P)"),st_numscalar("r(p_L)"),st_numscalar("r(p_Z)"))
}
foreach var in intlrmix oldratio share1 gini_market ehii credreg finreform {
	xtunitroot fisher d.`var' if country!="no" & country!="ie" ///
	& country!="es" & country!="at" & country!="dk" & country!="ch" & year<2008, dfuller lags(3)
	mata: p2`var'=(st_numscalar("r(p_P)"),st_numscalar("r(p_L)"),st_numscalar("r(p_Z)"))
}
mata
all=(p1dhPPPl\p1cPPPl\p1ydPPPl\p1ppl\p1spmixl\p1share1\p1gini_market\p1ehii\p1oldratio\p1credreg\p1finreform ///
\p2intlrmix\p2oldratio\p2share1\p2gini_market\p2ehii\p2credreg\p2finreform)
st_matrix("all",all)
end
cd "C:\Users\Rafael Wildauer\Dropbox\Diss\paper 2\results"
xml_tab all, save("unit root tests dif.xml") replace



*********************************************************************************************
********************************Section 4: Debt Function**************************************
*********************************************************************************************


xtreg d.(dhl ydl ppl spmixl share1 intlrmix oldratio credregl) if country!="ie" ///
& country!="es" & country!="at" & country!="dk" & country!="ch" & year<2008, fe robust

*create country specific time trends
xi i.time|cntry
reg d.(dhl ydl) ppl spmixl share1 intlrmix oldratio credregl dum* _ItimXcntr_* if country!="ie" ///
& country!="es" & country!="at" & country!="dk" & country!="ch" & year<2008, robust


xtreg d.(dhl ydl) ppl spmixl share1 intlrmix oldratio credregl if country!="ie" ///
& country!="es" & country!="at" & country!="dk" & country!="ch" & year<2008, fe robust
```

```
*standard growth regressions, pre crisis only
	local iter=0
	foreach cred in finreforml credregl {
		foreach ineq in share1 gini_market ehii {
			local iter=`iter'+1

			xtreg d.(dhl ydl ppl spmixl `ineq' intlrmix oldratio `cred') if country!="ie" ///
			& country!="es" & country!="at" & country!="dk" & country!="ch" & year<2008, fe robust
			estimates store debtgs`iter'

		}
	}
*create excel file
	estout debtgs* using "$out\tab_1_short.csv", ///
	cells(b(star fmt(3)) se(par('( )) fmt(2))) ///
	starlevels( * 0.1 ** 0.05 *** 0.01)  ///
	legend label ///
	varlabels(D.ydPPPl ln(YDPPP) ///
	D.ppl ln(PP) ///
	D.spmixl ln(SP) ///
	D.share1 Top1 ///
	D.gini_market gini ///
	D.ehii ehii ///
	LD.finreform finref_t-1 ///
	LD.credreg credreg_t-1 ///
	D.intlrmix i ///
	D.oldratio old) ///
	stats(N r2 r2_a F, fmt(0 2 2 0) ///
	label(N R-sqr adjR-sqr F-stat)) replace

*grwoth regressions with lagged PP
	local iter=0
	foreach cred in finreforml credregl {
		foreach ineq in share1 gini_market ehii {
			local iter=`iter'+1
```

```
				xtreg d.(dhl ydl l.ppl spmixl `ineq' intlrmix oldratio `cred') if country!="ie" ///
				& country!="es" & country!="at" & country!="dk" & country!="ch" & year<2008, fe robust
				estimates store debtqlag`iter'

			}
		}
*create excel file
		estout debtqlag* using "$out\tab_1_lag_PP.csv", ///
		cells(b(star fmt(3)) se(par('( )) fmt(2))) ///
		starlevels( * 0.1 ** 0.05 *** 0.01)  ///
		legend label ///
		varlabels(D.ydPPPl ln(YDPPP) ///
		D.ppl ln(PP) ///
		D.spmixl ln(SP) ///
		D.share1 Top1 ///
		D.gini_market gini ///
		D.ehii ehii ///
		LD.finreform finref_t-1 ///
		LD.credreg credreg_t-1 ///
		D.intlrmix i ///
		D.oldratio old) ///
		stats(N r2 r2_a F, fmt(0 2 2 0) ///
		label(N R-sqr adjR-sqr F-stat)) replace


*MG on first differenced sample, pre crisis
		*create differences firts
			foreach var in dhl ydl share1 gini_market ehii ppl spmixl intlrmix oldratio finreforml credregl  {
				bysort cntry (year): gen d`var'=`var'[_n]-`var'[_n-1]
			}
		*run first differenced regressions
			local iter=0
			foreach cred in finreforml credregl {
				foreach ineq in share1 gini_market ehii {

					local iter=`iter'+1
```

```
			xtmg ddhl dydl dppl dspmixl d`ineq' dintlrmix doldratio d`cred' if  country!="ie" ///
			& country!="es" & country!="at" & country!="dk" & country!="ch" & year<2008, full
			estimates store debtmg`iter'
			mat betas`iter'=e(betas)
			mat tbetas`iter'=e(tbetas)

		}
}
*combined individual country results
		mat debtMG=J(78,8,1)
		mat debtMGtstat=J(78,8,1)
		forvalues j=1(6)78 {
			local row=(`j'+5)/6
			disp `row'
			mat debtMG[`j',1]=betas1[`row',1..8]
			mat debtMG[`j'+1,1]=betas2[`row',1..8]
			mat debtMG[`j'+2,1]=betas3[`row',1..8]
			mat debtMG[`j'+3,1]=betas4[`row',1..8]
			mat debtMG[`j'+4,1]=betas5[`row',1..8]
			mat debtMG[`j'+5,1]=betas6[`row',1..8]
			mat debtMGtstat[`j',1]=tbetas1[`row',1..8]
			mat debtMGtstat[`j'+1,1]=tbetas2[`row',1..8]
			mat debtMGtstat[`j'+2,1]=tbetas3[`row',1..8]
			mat debtMGtstat[`j'+3,1]=tbetas4[`row',1..8]
			mat debtMGtstat[`j'+4,1]=tbetas5[`row',1..8]
			mat debtMGtstat[`j'+5,1]=tbetas6[`row',1..8]
		}
		mat colnames debtMG=yd Q PP SP R old CS const
		mat colnames debtMGtstat=yd Q PP SP R old CS const
		mat list debtMG
		mat list debtMGtstat
		cd "H:\Data\chapter 2\results"
		cd "C:\Users\Rafael Wildauer\Dropbox\Diss\paper 2\results"
		xml_tab debtMG, save("debt_MG.xml") replace
		xml_tab debtMGtstat, save("debt_MGtstat.xml") replace
*create excel file
		estout debtmg* using "$out\tab_1_mg.csv", ///
		cells(b(star fmt(3)) se(par('( )) fmt(2))) ///
```

```
				starlevels( * 0.1 ** 0.05 *** 0.01)  ///
				legend label ///
				varlabels(D.ydPPPl ln(YDPPP) ///
				D.ppl ln(PP) ///
				D.spmixl ln(SP) ///
				D.share1 Top1 ///
				D.gini_market gini ///
				D.ehii ehii ///
				LD.finreform finref_t-1 ///
				LD.credreg credreg_t-1 ///
				D.intlrmix i ///
				D.oldratio old) ///
				stats(N r2 r2_a F, fmt(0 2 2 0) ///
				label(N R-sqr adjR-sqr F-stat )) replace

*extract individual R2 and F-stat results
	xtmg ddhl dydl dshare1 dppl dspmixl dintlrmix doldratio dcredregl if  country!="ie" ///
	& country!="es" & country!="at" & country!="dk" & country!="ch", full

	estimates store debtmg`iter'
	mat betas`iter'=e(betas)
	mat tbetas`iter'=e(tbetas)



************PMG: baseline
*share
	xtpmg d.(dhl ydl ppl spmixl share1 intlrmix oldratio) if country!="at" & ///
	country!="ch" & country!="dk" & country!="es" & country!="ie" & year<2008, ///
	lr(l.dhl ydl ppl spmixl share1 intlrmix oldratio) ec(ec) replace pmg
	estimates store debtpmgshortloc1
	*sign adjustment: BE, DE, FI, FR, GB, JP, NL, NO, SE (slow: SE, JP, DE)
	*not included: AU, CA, IT, US
	*test unity income elasticity
		test [ec]ydl=1 //Null of unit elasticity is not rejected
	*test unit PP elasticity
		test [ec]ppl=1
*share, core group
```

```
		xtpmg d.(dhl ydl ppl spmixl share1 intlrmix oldratio) if (country=="be" | ///
		country=="fi" | country=="fr" | country=="gb" | ///
		country=="nl" | country=="no") & year<2008, ///
		lr(l.dhl ydl ppl spmixl share1 intlrmix oldratio) ec(ec) replace pmg
		estimates store debtpmgshortloc2
*gini
		xtpmg d.(dhl ydl ppl spmixl gini_market intlrmix oldratio) if country!="at" & ///
		country!="ch" & country!="dk" & country!="es" & country!="ie" & year<2008, ///
		lr(l.dhl ydl ppl spmixl gini_market intlrmix oldratio) ec(ec) replace pmg
		estimates store debtpmgshortloc3
		*sign adjustment: DE, FI, FR, GB, IT, JP, NL, NO, SE (slow: DE, IT, JP, SE)
		*not included: AU, BE (barely), CA, US
		*test unity income elasticity
			test [ec]ydl=1 //Null of unit elasticity is not rejected
		*test unit PP elasticity
			test [ec]ppl=1
*gini, core group
		xtpmg d.(dhl ydl ppl spmixl gini_market intlrmix oldratio) if (country=="be" | ///
		country=="fi" | country=="fr" | country=="gb" | ///
		country=="nl" | country=="no") & year<2008, ///
		lr(l.dhl ydl ppl spmixl gini_market intlrmix oldratio) ec(ec) replace pmg
		estimates store debtpmgshortloc4
*create excel file
		estout debtpmgshortloc* using "$out\debt_pmg_short.csv", ///
		cells(b(star fmt(3)) se(par('( )) fmt(2))) ///
		starlevels( * 0.1 ** 0.05 *** 0.01)  ///
		legend label ///
		varlabels(D.ydPPPl d.ln(YDPPP) ydPPPl ln(YDPPP) ///
		D.ppl d.ln(PP) ppl ln(PP) ///
		D.spmixl d.ln(SP) spmixl ln(SP) ///
		D.share1 d.Top1 share1 Top1 ///
		D.gini_market d.gini gini_market gini ///
		D.ehii d.ehii ehii ehii ///
		D.finreform d.finref finreform finref ///
		D.credreg d.credreg credreg credreg ///
		D.intlrmix d.i intlrmix i) ///
		stats(N, fmt(0 2 2 0 3 3 3) ///
		label(N )) replace
```

```
*****************************************************************************************
**********************Section 4B: Debt Function with interactions******************************
*****************************************************************************************

*define PP bubble dummy based on countries
*Those countries with increases in real PP of more than 80% between 1995 and 2006:
*AU, FI, FR, UK, NL, NO, SE, US
*Thus non bubble countries: BE, CA, DE, IT, JP

*generate country based dummy
	gen PPbubble=1 if country=="au" | country=="fi" | country=="fr" | country=="gb" | country=="nl" ///
	| country=="no" | country=="se" | country=="us"
	replace PPbubble=0 if country=="de" | country=="jp" | country=="be" | country=="ca" | country=="it"

*standard growth regressions: bubble countries
	local iter=0
	foreach cred in finreforml credregl {
		foreach ineq in share1 gini_market {
			local iter=`iter'+1

			xtreg d.(dhl ydl ppl spmixl `ineq' intlrmix oldratio `cred') if PPbubble==1 & year<2008, fe robust
			estimates store debtbub`iter'

		}
	}
*standard growth regressions: non-bubble countries
	local iter=4
	foreach cred in finreforml credregl {
		foreach ineq in share1 gini_market {
			local iter=`iter'+1

			xtreg d.(dhl ydl ppl spmixl `ineq' intlrmix oldratio `cred') if PPbubble==0 & year<2008, fe robust
			estimates store debtbub`iter'

		}
	}
```

```
*create excel file
	estout debtbub* using "$out\tab_1_bubble.csv", ///
	cells(b(star fmt(2)) se(par('( )) fmt(2))) ///
	starlevels( * 0.1 ** 0.05 *** 0.01)  ///
	legend label ///
	varlabels(D.ydPPPl ln(YDPPP) ///
	D.ppl ln(PP) ///
	D.spmixl ln(SP) ///
	D.share1 Top1 ///
	D.gini_market gini ///
	D.ehii ehii ///
	LD.finreform finref_t-1 ///
	LD.credreg credreg_t-1 ///
	D.intlrmix i ///
	D.oldratio old) ///
	stats(N r2 r2_a F, fmt(0 2 2 0) ///
	label(N R-sqr adjR-sqr F-stat)) replace


************PMG alternative PP bubble
*share: bubble
	xtpmg d.(dhl ydl share1 ppl spmixl intlrmix oldratio) if PPbubble==1 & year<2008, ///
	lr(l.dhl ydl share1 ppl spmixl intlrmix oldratio) ec(ec) replace pmg
	estimates store debtpmgaltPP1
	*AU and US no stat sign adjustment
	*test unity income elasticity
		test [ec]ydl=1
	*test unit PP elasticity
		test [ec]ppl=1
*gini: bubble
	xtpmg d.(dhl ydl gini_market ppl spmixl intlrmix oldratio) if (country=="au" | ///
	country=="fi" | country=="fr" | country=="gb" | ///
	country=="nl" | country=="no" | country=="se" | country=="us") & year<2008, ///
	lr(l.dhl ydl gini_market ppl spmixl intlrmix oldratio) ec(ec) replace pmg
	estimates store debtpmgaltPP2
	*AU, SE and US no stat sign adjustment
	*test unity income elasticity
		test [ec]ydl=1
```

```
	*test unit PP elasticity
		test [ec]ppl=1
*share: non-bubble
	xtpmg d.(dhl ydl share1 ppl spmixl intlrmix oldratio) if (country=="be" | ///
	country=="ca" | country=="de" | country=="it" | country=="jp") & year<2008, ///
	lr(l.dhl ydl share1 ppl spmixl intlrmix oldratio) ec(ec) replace pmg
	estimates store debtpmgaltPP3
	test [ec]ydl=1
	test [ec]ppl=1
	*CA and IT no stat sign adjustment
	*no stat sign adjustment
*gini: non-bubble
	xtpmg d.(dhl ydl gini_market ppl spmixl intlrmix oldratio) if (country=="be" | ///
	country=="ca" | country=="de" | country=="it" | country=="jp") & year<2008, ///
	lr(l.dhl ydl gini_market ppl spmixl intlrmix oldratio) ec(ec) replace pmg
	estimates store debtpmgaltPP4
	test [ec]ydl=1
	test [ec]ppl=1
	*BE and CA no stat sign adjustment
	*no stat sign adjustment
*create excel file
	estout debtpmgaltPP* using "$out\debt_pmg_altPPbubb.csv", ///
	cells(b(star fmt(3)) se(par('( )) fmt(2))) ///
	starlevels( * 0.1 ** 0.05 *** 0.01)  ///
	legend label ///
	varlabels(D.ydPPPl d.ln(YDPPP) ydPPPl ln(YDPPP) ///
	D.ppl d.ln(PP) ppl ln(PP) ///
	D.spmixl d.ln(SP) spmixl ln(SP) ///
	D.share1 d.Top1 share1 Top1 ///
	D.gini_market d.gini gini_market gini ///
	D.ehii d.ehii ehii ehii ///
	D.finreform d.finref finreform finref ///
	D.credreg d.credreg credreg credreg ///
	D.intlrmix d.i intlrmix i) ///
	stats(N, fmt(0 2 2 0 3 3 3) ///
	label(N )) replace
```

```
*****************************************************************************************
*************************Section 5: Consumption Function**********************************
*****************************************************************************************



*Fixed effects estimator with local currency
	local iter=0
	foreach cred in finreforml credregl {
		foreach ineq in share1 gini_market ehii {

			local iter=`iter'+1
			xtreg d.(cl ydl `ineq' ppl spmixl intlrmix oldratio `cred') if country!="ie" ///
			& country!="es" & country!="at" & country!="dk" & country!="ch", fe robust
			estimates store cbaseloc`iter'
		}
	}
*create Excel tables
	estout cbaseloc* using "$out\consumption fe loc.csv", ///
	cells(b(star fmt(3)) se(par('( )) fmt(2))) ///
	starlevels( * 0.1 ** 0.05 *** 0.01)  ///
	legend label ///
	varlabels(D.ydl ln(YD) ///
	D.ppl ln(PP) ///
	D.spmixl ln(SP) ///
	D.share1 Top1 ///
	D.gini_market gini ///
	D.ehii ehii ///
	D.intlrmix i) ///
	stats(N r2_a F, fmt(0 2 0) ///
	label(N adjR-sqr F-stat)) replace



*standard growth regressions: bubble countries
	local iter=0
	foreach cred in finreforml credregl {
		foreach ineq in share1 gini_market {
```

```
                local iter=`iter'+1
                xtreg d.(cl ydl `ineq' ppl spmixl intlrmix oldratio `cred') if PPbubble==1 & year<2008, fe robust
                estimates store consbub`iter'

            }
    }
*standard growth regressions: non-bubble countries
    local iter=4
    foreach cred in finreforml credregl {
            foreach ineq in share1 gini_market {
                local iter=`iter'+1
                xtreg d.(cl ydl `ineq' ppl spmixl intlrmix oldratio `cred') if PPbubble==0 & year<2008, fe robust
                estimates store consbub`iter'

            }
    }
*create excel file
    estout consbub* using "$out\cons_bubble.csv", ///
    cells(b(star fmt(3)) se(par('( )) fmt(2))) ///
    starlevels( * 0.1 ** 0.05 *** 0.01)  ///
    legend label ///
    varlabels(D.ydPPPl ln(YDPPP) ///
    D.ppl ln(PP) ///
    D.spmixl ln(SP) ///
    D.share1 Top1 ///
    D.gini_market gini ///
    D.ehii ehii ///
    LD.finreform finref_t-1 ///
    LD.credreg credreg_t-1 ///
    D.intlrmix i ///
    D.oldratio old) ///
    stats(N r2 r2_a F, fmt(0 2 2 0) ///
    label(N R-sqr adjR-sqr F-stat)) replace



*MG on first differenced sample, pre crisis
    *create differences firts
```

```
	foreach var in cl  {
		bysort cntry (year): gen d`var'=`var'[_n]-`var'[_n-1]
	}
*run first differenced regressions
	local iter=0
	foreach cred in finreforml credregl {
		foreach ineq in share1 gini_market ehii {

			local iter=`iter'+1

			xtmg dcl dydl d`ineq' dppl dspmixl dintlrmix doldratio d`cred' if  country!="ie" ///
			& country!="es" & country!="at" & country!="dk" & country!="ch" & year<2008, full
			estimates store consmg`iter'
			mat betas`iter'=e(betas)
			mat tbetas`iter'=e(tbetas)

		}
	}
	*combined individual country results
		mat consMG=J(78,8,1)
		mat consMGtstat=J(78,8,1)
		forvalues j=1(6)78 {
			local row=(`j'+5)/6
			disp `row'
			mat consMG[`j',1]=betas1[`row',1..8]
			mat consMG[`j'+1,1]=betas2[`row',1..8]
			mat consMG[`j'+2,1]=betas3[`row',1..8]
			mat consMG[`j'+3,1]=betas4[`row',1..8]
			mat consMG[`j'+4,1]=betas5[`row',1..8]
			mat consMG[`j'+5,1]=betas6[`row',1..8]
			mat consMGtstat[`j',1]=tbetas1[`row',1..8]
			mat consMGtstat[`j'+1,1]=tbetas2[`row',1..8]
			mat consMGtstat[`j'+2,1]=tbetas3[`row',1..8]
			mat consMGtstat[`j'+3,1]=tbetas4[`row',1..8]
			mat consMGtstat[`j'+4,1]=tbetas5[`row',1..8]
			mat consMGtstat[`j'+5,1]=tbetas6[`row',1..8]
		}
		mat colnames consMG=yd Q PP SP R old CS const
```

```
            mat colnames consMGtstat=yd Q PP SP R old CS const
            mat list consMG
            mat list consMGtstat
            cd "H:\Data\chapter 2\results"
            cd "C:\Users\Rafael Wildauer\Dropbox\Diss\paper 2\results"
            xml_tab consMG, save("cons_MG.xml") replace
            xml_tab consMGtstat, save("cons_MGtstat.xml") replace
        *create excel file
            estout consmg* using "$out\cons_mg.csv", ///
            cells(b(star fmt(3)) se(par('( )) fmt(2))) ///
            starlevels( * 0.1 ** 0.05 *** 0.01)  ///
            legend label ///
            varlabels(D.ydl ln(YD) ///
            D.ppl ln(PP) ///
            D.spmixl ln(SP) ///
            D.share1 Top1 ///
            D.gini_market gini ///
            D.ehii ehii ///
            LD.finreform finref_t-1 ///
            LD.credreg credreg_t-1 ///
            D.intlrmix i ///
            D.oldratio old) ///
            stats(N r2 r2_a F, fmt(0 2 2 0) ///
            label(N R-sqr adjR-sqr F-stat )) replace
*simplified MG without distribution
    xtmg dcl dydl dppl dspmixl dintlrmix doldratio dcredregl if  country!="ie" ///
    & country!="es" & country!="at" & country!="dk" & country!="ch" & year<2008, full
```

Bibliography

Abiad, A., Detragiache, E. & Tressel, T., 2008. A New Database of Financial Reforms. *IMF Working Paper No 266*.

Alesina, A. & Rodrik, D., 1994. Distributive Politics and Economic Growth. *The Quarterly Journal of Economics*, 109(2), pp.465–490.

Alpizar, F., Carlsson, F. & Johansson-Stenman, O., 2005. How much do we care about absolute versus relative income and consumption? *Journal of Economic Behavior & Organization*, 56(3), pp.405–421.

Alvarez-Cuadrado, F. et al., 2012. Envy and habits: Panel data estimates of interdependent preferences. *Banco De Espana Documentos de Trabajo No 1213*.

Alvarez-Cuadrado, F. & Vilalta, M.E.-A., 2012. Income inequality and saving. *IZA Discussion Paper No 7083*.

Anundsen, A.K. & Jansen, E.S., 2013. Self-reinforcing effects between housing prices and credit. *Journal of Housing Economics*, 22(3), pp.192–212.

Arestis, P. & González, A.R., 2014. Bank credit and the housing market in OECD countries. *Journal of Post Keynesian Economics*, 36(3), pp.467–490.

Arestis, P. & González, A.R., 2013. Endogenous Bank Credit and Its Link to Housing in OECD Countries. *Levy Economics Institute Working Paper No 750*.

Aron, J. et al., 2012. Credit, Housing Collateral, and Consumption: Evidence From Japan, the U.K., and the U.S. *Review of Income and Wealth*, 58(3), pp.397–423.

Atkinson, A.B., Piketty, T. & Saez, E., 2011. Top Incomes in the Long Run of History. *Journal of Economic Literature*, 49(1), pp.3–71.

Attanasio, O.P. et al., 2009. Booms and Busts: Consumption, House Prices and Expectations. *Economica*, 76(301), pp.20–50.

Attanasio, O.P. & Weber, G., 2010. Consumption and Saving: Models of Intertemporal Allocation and Their Implications for Public Policy. *Journal of Economic Literature*, 48(3), pp.693–751.

Attanasio, O.P. & Weber, G., 1994. The UK consumption boom of the late 1980s: Aggregate implications of microeconomic evidence. *The Economic Journal*, 104(November), pp.1269–1302.

Barba, A. & Pivetti, M., 2009. Rising household debt: its causes and macroeconomic implications-a long-period analysis. *Cambridge Journal of Economics*, 33(1), pp.113–138.

Barkbu, B. et al., 2015. *Investment in the Euro Area: Why Has It Been Weak?*,

Behringer, J. & Treeck, T. van, 2013. Income distribution and current account: A sectoral perspective. *IMK Working Paper No 125*, (November).

Belabed, C.A., Theobald, T. & van Treeck, T., 2013. *Income distribution and current account imbalances*, IMK Working Paper 126.

Bernanke, B., Gertler, M. & Gilchrist, S., 1998. The Financial Accelerator in a Quantitative Business Cycle Framework. *NBER working paper*, (6455).

Bertrand, M. & Morse, A., 2013. Trickle-Down Consumption. *NBER Working Paper No 18883*.

Bezemer, D., Grydaki, M. & Zhang, L., 2014. Is Financial Development Bad for Growth? *SOM Research Reports*, 14016–GEM.

Bezemer, D. & Zhang, L., 2014. From Boom to Bust in the Credit Cycle: The Role of Mortgage Credit. *SOM Research Reports*, 14025–GEM.

Bhattacharya, S. et al., 2015. A Reconsideration of Minsky's Financial Instability Hypothesis. *Journal of Money, Credit and Banking*, 47(5), pp.931–973.

Blackburne, E.F. & Frank, M.W., 2007. Estimation of nonstationary heterogeneous panels. *The Stata Journal*, 7(2), pp.197–208.

Blanchard, O., 2009. The Crisis: Basic Mechanisms, and Appropriate Policies. *IMF Working Paper No 09/80*.

Bordo, M.D. & Meissner, C.M., 2012. Does inequality lead to a financial crisis? *Journal of International Money and Finance*, 31(8), pp.2147–2161.

Borio, C., 2014. The financial cycle and macroeconomics: What have we learnt? *Journal of Banking and Finance*, 45(1), pp.182–198.

Bostic, R., Gabriel, S. & Painter, G., 2009. Housing wealth, financial wealth, and consumption: New evidence from micro data. *Regional Science and Urban Economics*, 39(1), pp.79–89.

Brancaccio, E. & Fontana, G., 2013. "Solvency rule" versus "taylor rule": An alternative interpretation of the relation between monetary policy and the economic crisis. *Cambridge Journal of Economics*, 37(1), pp.17–33.

Brown, C., 1997. Consumer credit and the propensity to consume: evidence from 1930. *Journal of Post Keynesian Economics*, 19(4).

Brown, C., 2007. Financial engineering, consumer credit, and the stability of effective demand. *Journal of Post Keynesian Economics*, 29(3), pp.427–450.

Campbell, J.Y. & Cocco, J.F., 2007. How do house prices affect consumption? Evidence from micro data. *Journal of Monetary Economics*, 54(3), pp.591–621.

Cardaci, A., 2014. Inequality , Household Debt and Financial Instability : An Agent-Based Perspective.

Carr, M.D. & Jayadev, A., 2015. Relative Income and Indebtedness: Evidence from Panel Data. *Review of Income and Wealth*, 61(4), pp.759–772.

Carroll, C.D., 2001. A Theory of the Consumption Function, With and Without Liquidity Constraints. *Journal of Economic Perspectives*, 15(3), pp.23–45.

Charles, S., 2008. A Post-Keynesian Model of Accumulation with a Minskyan Financial Structure. *Review of Political Economy*, 20(3), pp.319–331.

Choi, I., 2001. Unit root tests for panel data. *Journal of International Money and Finance*, 20(2), pp.249–272.

Christen, M. & Morgan, R.M., 2005. Keeping Up With the Joneses: Analyzing the Effect of Income Inequality on Consumer Borrowing. *Quantitative Marketing and Economics*, 3(2), pp.145–173.

Chrystal, K.A. & Mizen, P., 2005. A Dynamic Model of Money, Credit, and Consumption: A Joint Model for the UK Household Sector. *Journal of Money, Credit and Banking*, 37(1), pp.119–143.

Coibion, O. et al., 2016. Does Greater Inequality Lead to More Household Borrowing? New Evidence from Household Data. *Federal Reserve Bank of San Francisco Working Paper Series*, (20).

Cooper, D., 2012. House Price Fluctuations: The Role of Housing Wealth as Borrowing Collateral. *Review of Economics and Statistics*, 95(4), pp.1183–1197.

Cooper, D. & Dynan, K., 2016. Wealth Effects and Macroeconomic Dynamics. *Journal of Economic Surveys*, 30(1), pp.34–55.

Cynamon, B.Z. & Fazzari, S.M., 2016. Inequality, the Great Recession and slow recovery. *Cambridge Journal of Economics*, 40(2), pp.373–399.

Dell'Ariccia, G., Igan, D. & Laeven, L., 2012. Credit Booms and Lending Standards: Evidance from the Subprime Mortgage Market. *Journal of Money, Credit and Banking*, 44(2), pp.367–384.

Drechsel-Grau, M. & Schmid, K.D., 2014. Savings decisions under upward-looking comparisons. *Journal of Economic Behavior and Organization*, 106, pp.254–268.

Duesenberry, J.S., 1949. *Income, Saving and the Theory of Consumer Behavior*, Harvard University Press.

Dynan, K., 2010. Wealth Effects and the Changing Economy. *manuscript*.

Dynan, K.E. & Kohn, D.L., 2007. The Rise in U.S. Household Indebtedness: Causes and Consequences. *Finance and Economics Discussion Series*, 37.

Dynan, K.E., Skinner, J. & Zeldes, S.P., 2004. Do the Rich Save More? *Journal of Political Economy*, 112(2), pp.397–444.

Eberhardt, M., 2012. Estimating panel time-series models with heterogeneous slopes. *The Stata Journal*, 12(1), pp.61–71.

ECB, 2013. *The Eurosystem Household Finance and Consumption Survey Methodological Report for the First Wave*, Frankfurt.

Eckerstorfer, P. et al., 2016. Correcting for the Missing Rich: An Application to Wealth Survey Data. *Review of Income and Wealth*, 62(4), pp.605–627.

Eckerstorfer, P. & Wendner, R., 2013. Asymmetric and non-atmospheric consumption externalities, and efficient consumption taxation. *Journal of Public Economics*, 106, pp.42–56.

Égert, B., Backé, P. & Zumer, T., 2006. Credit growth in central and eastern Europe. *ECB working paper No 687*.

Eichengreen, B. & Mitchener, K., 2003. The Great Depression as a Credit Boom Gone Wrong. *BIS Working Papers*, (137).

European Commission, 2015. Investment dynamics in the euro area since the crisis. *Quarterly Report on the Euro Area*, 14(1), pp.35–43.

Fernandez-Corugedo, E. & Muellbauer, J., 2006. Consumer credit conditions in the United Kingdom. *Bank of England Working Paper no 314*, (314).

Fisher, I., 1933. The Debt-Deflation Theory of Great Depressions. *Econometrica*, 1(4), pp.337–357.

Frank, R., 1985. The demand for unobservable and other nonpositional goods. *The American Economic Review*, 75(1), pp.101–116.

Frank, R.H., Levine, A.S. & Dijk, O., 2014. Expenditure Cascades. *Review of Behavioral Economics*, 1(1–2), pp.55–73.

Frey, W.H., 2015. Census shows modest declines in black-white segregation. *The Avenue*. Available at: https://www.brookings.edu/blog/the-avenue/2015/12/08/census-shows-modest-declines-in-black-white-segregation/ [Accessed November 24, 2016].

Friedline, T., Masa, R.D. & Chowa, G.A.N., 2015. Transforming wealth: Using the inverse hyperbolic sine (IHS) and splines to predict youth's math achievement. *Social Science Research*, 49, pp.264–287.

Gimeno, R. & Martinez-Carrascal, C., 2010. The relationship between house prices and house purchase loans: The Spanish case. *Journal of Banking and Finance*, 34(8), pp.1849–1855.

Glaeser, E. & Vigdor, J., 2012. The end of the segregated century: Racial separation in America's neighborhoods, 1890-2012. *Civic Report*, 66, pp.1–36.

Godley, W. & Lavoie, M., 2007. *Monetary Economics*, New York: Palgrave Macmillan.

Goodhart, C. & Hofmann, B., 2008. House prices, money, credit, and the macroeconomy. *Oxford Review of Economic Policy*, 24(1), pp.180–205.

Greenspan, A. & Kennedy, J., 2008. Sources and uses of equity extracted from homes. *Oxford Review of Economic Policy*, 24(1), pp.120–144.

Gu, X. & Huang, B., 2014. Does inequality lead to a financial crisis? Revisited. *Review of Development Economics*, 18(3), pp.502–516.

Haurin, D.R. & Rosenthal, S.S., 2006. House Price Appreciation, Savings, and Consumer Expenditures. *manuscript*.

Heffetz, O., 2011. A test of conspicuous consumption: Visibility and income elasticities. *Review of Economics and Statistics*, 93(4), pp.1101–1117.

Hein, E. & Vogel, L., 2008. Distribution and growth reconsidered: empirical results for six OECD countries. *Cambridge Journal of Economics*, 32(3), pp.479–511.

Hendry, D.F., 1995. *Dynamic Econometrics*, Oxford University Press.

Hofmann, B., 2004. The Determinants of Bank Credit in Industrialized Countries: Do Property Prices Matter? *International Finance*, 7(2), pp.203–234.

Holt, R.P.F. & Greenwood, D.T., 2012. Negative Trickle-Down and the Financial Crisis of 2008. *Journal of Economic Issues*, 46(2), pp.363–370.

Horn, G. et al., 2009. From the financial crisis to the world economic crisis. *IMK Policy Brief*.

Hurst, E. & Stafford, F., 2004. Home is where the equity is: Liquidity constraints, refinancing and consumption. *Journal of Money, Credit and Banking*, 36(6), pp.985–1014.

Jappelli, T. & Pagano, M., 1994. Saving, Growth, and Liquidity Constraints. *Quarterly Journal of Economics*, 109(1), pp.83–109.

Juster, F.T. et al., 2005. The Decline in Household Saving and the Wealth Effect. *The Review of Economics and Statistics*, 87(4), pp.20–27.

Kapeller, J. & Schütz, B., 2014. Debt, boom, bust: a theory of Minsky-Veblen cycles. *Journal of Post Keynesian Economics*, 36(4), pp.781–814.

Keen, S., 2013. A monetary Minsky model of the Great Moderation and the Great Recession. *Journal of Economic Behavior and Organization*, 86, pp.221–235.

Kennickell, A., 2005. The Good Shepherd: Sample Design and Control for Wealth Measurement in the Survey of Consumer Finances. *Luxembourg Wealth Study Conference*.

Kennickell, A.B., 1998. Multiple imputation in the Survey of Consumer Finances. *Joint Statistical Meetings*.

Kennickell, A.B., 2008. The Role of Over-sampling of the Wealthy in the Survey of Consumer Finances. *IFC Bulletin*, (28), pp.403–408.

Kennickell, A.B., 2000. Wealth Measurement in the Survey of Consumer Finances: Methodology and Directions for Future Research. *Finance & Economic Discussion Paper, Federal Reserve Board Washington D.C.*

Kennickell, A.B. & McManus, D. a, 1993. Sampling for Household Financial Characteristics Using Frame Information on Past Income. *Proceedings of the Survey Research Methods Section, American Statistical Association*, 1, pp.88–97.

Kiff, J. & Mills, P., 2007. Money for nothing and checks for free: Recent developments in US subprime mortgage markets. *International Monetary Fund Working Paper*, pp.2–20.

Klein, M., 2015. Inequality and household debt : a panel cointegration analysis. *Empirica*, 42, pp.391–412.

Koo, R., 2011. The world in balance sheet recession: causes, cure, and politics. *Real World Economics Review*, (58), pp.19–37.

Kumhof, M. et al., 2012. Income Inequality and Current Account Imbalances. *IMF Working Paper 12/08*.

Kumhof, M. & Ranciere, R., 2010. Inequality, leverage and crises. *IMF Working Paper 10/268*.

Leamer, E.E., 2007. Housing is the Business Cycle. In *Economic Policy Symposium - Jackson Hole*. pp. 149–233.

Lehnert, A., 2004. Housing, Consumption, and Credit Constraints. *manuscript*.

Leigh, A. & Posso, A., 2009. TOP INCOMES AND NATIONAL SAVINGS. *Review of Income and Wealth*, 55(1), pp.57–74.

Lewis, C. et al., 2014. Investment Gaps after the Crisis. *OECD Economics Department Working Paper 1168*.

Li, H. & Zou, H., 2004. Savings and income distribution. *Annals of Economics and Finance*, 270, pp.245–270.

Ludwig, A. & Sløk, T., 2004. The Relationship between Stock Prices, House Prices and Consumption in OECD Countries. *Topics in Macroeconomics*, 4(1).

MacKinnon, J.G. & Magee, L., 1990. Transforming the Dependent Variable in Regression Models. *International Economic Review*, 31(2), pp.315–339.

Malinen, T., 2014. Does income inequality contribute to credit cycles? *MPRA Paper No. 52831*.

Maurer, J. & Meier, A., 2008. Smooth it Like the "Joneses"? Estimating Peer-Group Effects in Intertemporal Consumption Choice. *The Economic Journal*, 118(2006), pp.454–476.

Meng, X., Hoang, N.T. & Siriwardana, M., 2013. The determinants of Australian household debt: A macro level study. *Journal of Asian Economics*, 29, pp.80–90.

Mian, A. & Sufi, A., 2011. House prices, home equity-based borrowing, and the US household leverage crisis. *American Economic Review*, 101(5), pp.2132–2156.

Mian, A. & Sufi, A., 2009. The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis. *Quarterly Journal of Economics*, 124(4), pp.1449–1496.

Minsky, H.P., 1978. The Financial Instability Hypothesis: A Restatement. *Hyman P. Minsky Archive Paper 180*.

Muellbauer, J. & Murphy, A., 2008. Housing markets and the economy: The assessment. *Oxford Review of Economic Policy*, 24(1), pp.1–33.

Nikiforos, M., 2015. A Nonbehavioral Theory of Saving. *Levy Economics Institute Working Paper No. 844*.

Oikarinen, E., 2009. Interaction between housing prices and household borrowing: The Finnish case. *Journal of Banking and Finance*, 33(4), pp.747–756.

Onaran, Ö., Stockhammer, E. & Grafl, L., 2011. Financialisation, income distribution and aggregate demand in the USA. *Cambridge Journal of Economics*, 35(4), pp.637–661.

Paiella, M., 2009. The Stock Market, Housing and Consumer Spending: a Survey of the Evidence on Wealth Effects. *Journal of Economic Surveys*, 23(5), pp.947–973.

Palley, T.I., 2010. The Limits of Minsky's Financial Instability Hypothesis as an Explanation of the Crisis. *Monthly Review*, (October), pp.28–44.

Perugini, C., Holscher, J. & Collie, S., 2016. Inequality, credit and financial crises. *Cambridge Journal of Economics*, 40(1), pp.227–257.

Pesaran, H.M., Shin, Y. & Smith, R.P., 1999. Pooled Mean Group Estimation of Dynamic Heterogeneous Panels. *Journal of the American Statistical Association*, 94(446), pp.621–634.

Pesaran, H.M. & Smith, R., 1995. Estimating long-run relationships from dynamic heterogeneous panels. *Journal of Econometrics*, 68(1), pp.79–113.

Piketty, T., 2014. *Capital in the Twenty-First Century*, Cambridge: Harvard University Press.

Purnanandam, A., 2011. Originate-to-distribute model and the subprime mortgage crisis. *Review of Financial Studies*, 24(6), pp.1881–1915.

Rajan, R.G., 2010. *Fault Lines: How Hidden Fractures Still Threaten the World Economy*, Princeton: Princeton University Press.

Ravina, E., 2007. Habit Persistence and Keeping Up with the Joneses: Evidence from Micro Data. *manuscript*.

Ricks, M., 2016. *The Money Problem*, Chicago and London: The University of Chicago Press.

Romer, D., 2012. *Advanced Macroeconomics* 4th ed., New York: McGraw-Hill.

Rubaszek, M. & Serwa, D., 2014. Determinants of credit to households: An approach using the life-cycle model. *Economic Systems*, 38(4), pp.572–587.

Rubin, D., 1987. *Multiple Imputations for Nonresponse in Surveys*, New York: Wiley & Sons.

Ryoo, S., 2015. Household debt and housing bubble: A Minskian approach to boom-bust cycles. *UMAS Amherst Department of Economics Working Paper 2015-08*.

Ryoo, S. & Kim, Y.K., 2014. Income Distribution, Consumer Debt and Keeping up with the Joneses. *Metroeconomica*, 65(4), pp.585–618.

Salotti, S., 2012. Wealth effects in the US: Evidence from the combination of two surveys. *The Economic and Social Review*, 43(1), pp.67–98.

Dos Santos, C.H., 2005. A stock-flow consistent general framework for formal Minskyan analyses of closed economies. *Journal of Post Keynesian Economics*, 27(4), pp.711–735.

Schmidt-Hebbel, K. & Servén, L., 2000. Does income inequality raise aggregate saving? *Journal of Development Economics*, 61(2), pp.417–446.

Schularick, M. & Taylor, A.M., 2012. Credit booms gone bust: Monetary policy, leverage cycles, and financial crises, 1870-2008. *American Economic Review*, 102(2), pp.1029–1061.

Singer, E., 2006. Nonresponse Bias in Household Surveys. *Public Opinion Quarterly*, 70(5), pp.637–645.

Smith, D., 2001. International evidence on how income inequality and credit market imperfections affect private saving rates. *Journal of Development Economics*, 64(1), pp.103–127.

Stockhammer, E., 2009. The finance-dominated accumulation regime, income distribution and the present crisis. *Papeles de Europa*, 19, pp.58–81.

Stockhammer, E. & Wildauer, R., 2016. Debt-driven growth? Wealth, distribution and demand in OECD Countries. *Cambridge Journal of Economics*, 40(6), pp.1609–1634.

Taylor, J.B., 2009. The Financial Crisis and the Policy Responses: An Empirical Analysis of What Went Wrong. *NBER Working Paper No 14631*, pp.1–30.

The Volcker Alliance, 2016. *Unfinished business*, New York.

Tobin, J., 1969. Equilibrium Approach To Monetary Theory. *Journal of Money, Credit and Banking*, 1(1), pp.15–29.

Tobin, J., 1982. Money and Finance in the Macroeconoic Process. *Journal of Money, Credit, and Banking*, 14(2), pp.171–204.

Van Treeck, T., 2014. Did inequality cause the U.S. financial crisis? *Journal of Economic Surveys*, 28(3), pp.421–448.

Veblen, T., 1899. *The Theory of the Leisure Class*, Oxford University Press.

Vermeulen, P., 2014. How fat is the top tail of the wealth distribution? *ECB working paper No 1692*.

Wildauer, R., 2016. Determinants of US Household Debt: New Evidence from the SCF. *PKSG working paper 1608*.

Zezza, G., 2008. U.S. growth, the housing market, and the distribution of income. *Journal of Post Keynesian Economics*, 30(3), pp.375–401.

Zucman, G., 2015. *The Hidden Wealth of Nations*, University of Chicago Press.