Even Strøm Engebretsen and Daniel Johansson

"Cash-out" refinancing – funding entry into self-employment?

A study on self-employed using a matching approach

Master's thesis in Financial Economics

Trondheim, June 2015

Supervisor: Dr. Xunhua Su

Norwegian University of Science and Technology
Faculty of Social Sciences and Technology Management
Department of Economics

NTNU – Trondheim
Norwegian University of
Science and Technology

Abstract

In recent years, the Norwegian housing prices and household debt to disposable income have reached extraordinary new heights. Su et. al (2014) points to cash-out refinancing as one of the key drivers behind the debt to income ratios. As a contribution to more knowledge about the real effects of cash-out refinancing, the present thesis studies a causal relationship between entry into self-employment and cash-out. The study has two main findings. First, a sizeable amount of the total cash-out appears to be an effect of the entry into self-employment, indicating that cash-out refinancing is to some degree used as funding for starting businesses. Second, cash-out for funding entry into self-employment is substantially higher for women, supporting previous research showing that women attracts less start-up capital than men. Our results contrasts previous findings where cash-out has first and foremost been related to financial difficulties, which may have potential destabilizing consequences (Mian and Sufi, 2011). Our study indicates that there are still some unexplained effects regarding cash-out. Hence, our findings call for more research on what the cash-out is used for, to further enlighten the real effects to the national economy.

Keywords: *Household debt, cash-out, self-employment, propensity score matching, gender difference*

Preface

This thesis is submitted in partial fulfillment of the requirements for the degree MSc. at the Norwegian University of Science and Technology. The master thesis is carried out as a cooperation between Even Strøm Engebretsen and Daniel Johansson. The data applied in the present thesis are retrieved from Statistics Norway's (SSB) EU-SILC survey (European Union Statistics on Income and Living Conditions) from 2012. The data set has been anonymized and made available by the Norwegian Social Science Data Service (NSD). Neither SSB nor NSD are responsible for the analyses or the interpretations of the data presented in this thesis. Thanks and appreciation to Ingvild A. Krogh at NSD for providing us with the data needed.

Further, we would like to thank our supervisor Dr. Xunhua Su for helpful guidance and valuable comments during our work with this thesis.

Trondheim, June, 2015

Even Strøm Engebretsen and Daniel Johansson

Table of contents

List of figures

List of tables

Abbreviations

Acronyms that need explanations:

ATE Average Treatment Effect

ATT Average Treatment effect on Treated

ATU Average Treatment effect on Untreated

CIA Conditional Independence Assumption

NN Nearest Neighbors

PSM Propensity Score Matching

1 Introduction

In the last couple of decades, Norway has seen an ever-growing household debt and house prices reaching new heights. The debt to disposable income ratio reached 227% at the end of 2014, and in response to these high levels and the continuous growth, there has been much attention to the Norwegian housing market in the media, and both politicians and researchers questions the sustainability of the household debt. Recently, this led the Norwegian Ministry of Finance to request the Financial Supervisory Authorities of Norway to provide solutions for dampening the surging house prices (Dagens Næringsliv, 2015).

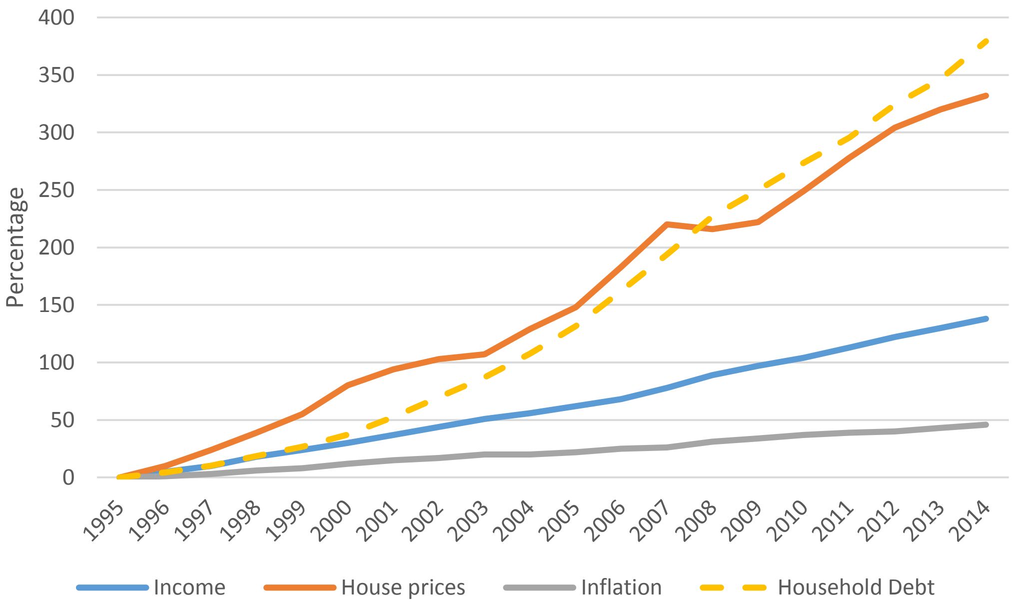

Figure 1.1: Household debt, household income, inflation and house prices, percentage increase since 1995
Source: Statistics Norway

With the current Norwegian debt situation, Su et al. (2014) recognizes the need for investigation of home-equity based borrowing in the Norwegian market. Based on data from the Survey on Income and Living Conditions (Statistics Norway, 2012), they find that home-equity based borrowing, which they refer to as cash-out, constitutes a substantial part of the high debt-to-income ratio. They also finds that cash-out relates to financial difficulties, which could be worrisome for the stability of the Norwegian economy.

As said by Su et al. (2014), their findings call for more research on cash-out in the Norwegian housing market. It is important to ask what the real effects from cash-out are, dependent on how households spend their newly borrowed money. Research up to this point have not found any conclusive evidence on how the cashed-out money is spent. Our contribution to reduce this uncertainty is to investigate whether cash-out refinancing are used as funding for small businesses and start-ups. We try to measure this by looking at individuals in the transition from wage-employment to self-employment. This is an important question to ask, as the cash-out effect has so far mostly been studied in negative contexts in relation to financial difficulty of the household and macro-economic instability caused by high debt-to-income ratio. However, for a more complete picture of the real effects from cash-out, we analyze if cash-out is used to fund new businesses, which puts the effects from cash-out in a more positive perspective. As Villund (2005) points out, even though self-employment constitutes a small part of the labor force, there is much interest in self-employment in respect to entrepreneurship, job creation and immigrants participation in the labor market. Especially for women, there seems to be an untapped entrepreneurial potential which could contribute to innovation and economic growth (OECD, 2004).

In our analysis, we raise an interesting question about a causal relationship between the self-selection into self-employment and the level of cash-out refinancing. Limited availability of funding and restricted access to credit can set a stopper for the establishment of new start-ups and small businesses, hence our idea is to investigate whether cash-out is used as a way to conquer this problem. When comparing raw data, the estimation of the difference in the level of cash-out between self-employed and wage-employed individuals give rise to possible bias, and to overcome this problem we have chosen to use a matching method. To address this issue of self-selection bias caused by non-random differences between self-employed and wage-employed individuals, we use propensity score matching (PSM). The PSM approach allows for construction of counterfactual outcomes, which is the level of cash-out the self-employed individuals would have had if they had not become self-employed. In such way, the PSM approach can measure causal effects.

In our analysis, the results from matching give an indication of a positive amount of cash-out caused by the self-selection into self-employment, indicating that home-equity extraction is used as a way of funding start-ups. Even though the matching procedure indicates a positive effect, we question the robustness of the results and further analyze a subsample where we exclude lower densities of estimated propensity scores. This procedure provides higher robustness, which supports our hypothesis. We have also addressed the issue of unobserved heterogeneity by conducting analyses of how sensitive the estimated results are to unobserved bias.

With the prevalent low participation of women engaging in self-employment, we further discuss the possibility of different cash-out tendencies among women and men who are self-employed. By the same matching procedure we find that women who becomes self-employed cashes-out considerably more than men. This is an interesting finding in relation to how women are more liquidity constrained when they want to start a business (Rønsen, 2012).

This study aims to contribute to more research on who cashes out and why. We have chosen to look at cash-out in relation to self-employment, by using the cash-out model developed by Su et al. (2014), which puts our thesis as an extension of their paper and among studies on self-employment, e.g. Rønsen (2012) or Disney and Gathergood (2009). We try to figure out if the choice of becoming self-employment is significant positively related to the level of cash-out when compared to other similar individuals. The matching procedure, indicates that the act of becoming self-employed are a cause for cash-out, and constitutes a sizeable amount of the total cash-out within the sample.

The structure of this thesis will be as follows: in chapter 2 we will cover relevant background literature of self-employment, house prices and debt, and cash-out refinancing. In chapter 3 we will formalize our research goals and chapter 4 will cover data material and descriptive statistics. Chapter 5 will cover theory of our empirical framework, and modeling of cash-out. Our main analysis is covered in chapter 6, with some further questions in chapter 7. Chapter 8 will conclude our thesis with discussion of the main results from chapter 6 and 7.

2 Background and literature

2.1 Characteristics on self-employed

Self-employment has recently become a popular area of research, and there is quite a few studies in various contexts, among them housing wealth and liquidity constraints (Hurst and Lusardi, 2004, Disney and Gathergood, 2009), wages and income differences (Praag and Raknerud, 2014, Kaiser and Malchow-Møller, 2011), innovation and economic growth (OECD, 2004), gender differences (Rønsen, 2012) and several descriptive statistical reports (Stambøl, 2008, Fjærli et al., 2013), all with self-employment as the main perspective. Self-employment as a topic can also be related to the housing market literature, as we investigate differences in the amount of home-equity refinancing in the Norwegian market, see e.g. Mian and Sufi (2011) or Su et al. (2014).

2.1.1 Research and statistical properties

Between 1996 and 2014, the average of self-employed individuals as percentage of the total labor stock was about 7.3%, and quite steady the whole time, illustrated in figure 2.1.

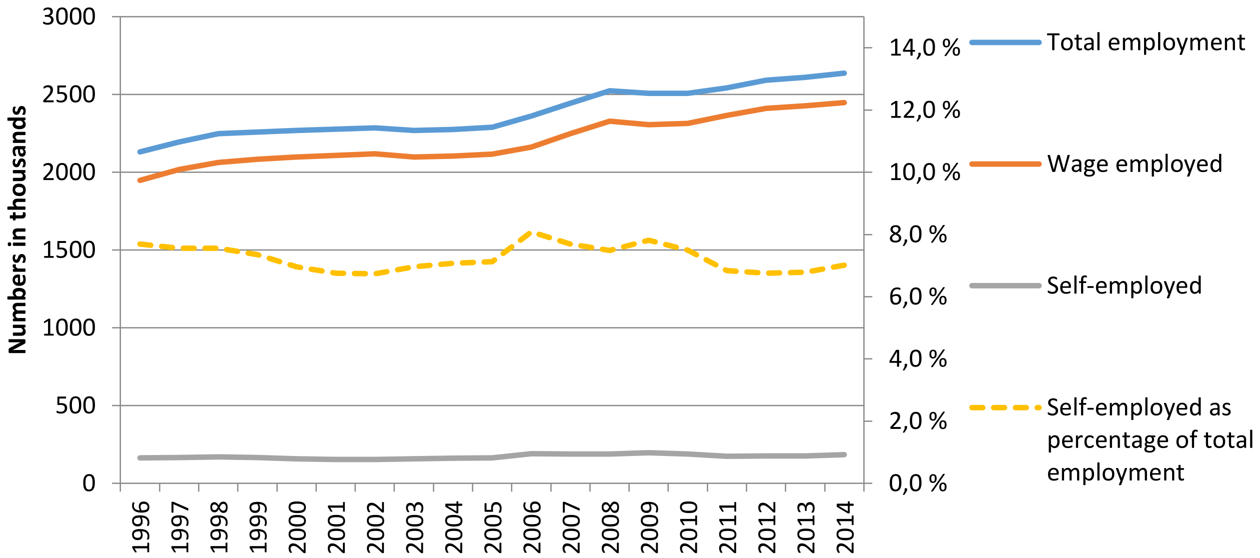

Figure 2.1: Total employment, wage employed and self-employed and self-employed as percentage of total employment
Source: Statistics Norway, Arbeidskraftundersøkelsen

Villund (2005) studies distribution and mobility of self-employed individuals in Norway, using data from 1997 to 2003. He finds that in general, significantly more men than women are self-employed, the average age is higher and the level of education is substantially lower than regular wage-employed individuals. He also points out that the proportion of self-employed are higher in rural than in central areas. This relates to the high amount (more than 50%[1]) of self-employed individuals that constitutes the primary sector (mostly farmers and fishermen) and craftsmen. The propensity to become self-employed is also higher in regions with a tighter job market and higher unemployment and most of the recruitment into self-employment comes from people outside the workforce. These demographical characteristics are also in accordance with Stambøl (2008), which uses data from the Labor force survey conducted by Statistics Norway.

Stambøl (2010) has also studied to what extent self-employed individuals sustain their business over time. More than 20% shut down their business after only one year, and in a five-year period, only 50% survived. Self-employed men last longer than women, and those who start a new business for their first time, are more likely to quit sooner. Middle-aged individuals with medium term education are most likely to have success with their business, and are less likely to give up. Many of those who end their career as self-employed, either become wage-employed or leave the workforce. Kaiser and Malchow-Møller (2011) used data on Danish men to quantify effects of past self-employment on subsequent earnings. Their econometric approach, which is the same as ours (PSM[2]), on average suggest that a period of self-employment actually lower hourly wages in their subsequent jobs. However, their findings show that hourly wages may increase, if they become wage employed in the same sector as they were self-employed. Hence, the negative effect on wages might be due to sector changes and not necessarily self-employment directly.

Rønsen (2012) studies the puzzle of the low women to men ratio of the self-employed individuals. She points to previous studies which show that female entrepreneurs attracts less capital and start businesses with scarce financial resources compared to men. In accordance with Rønsen, OECD (2004) also finds that women's access to capital is more restricted than men's. The likelihood for newly started businesses to survive early growth phases have been found to be equal for both men and women in Norway, but high growth companies are primarily started, owned and run by men (Ljunggren, 2008). A psychological perspective of the self-

[1] According to Villund (2005).
[2] More on this in Chapter 5

employed are contributed by e.g. Brown et al. (2011) and Ekelund et al. (2005), who studies why some individuals chose to become self-employed instead of being wage employed. They find that individuals with less risk-averse properties have a larger probability of future self-employment.

Rønsen (2012) further discuss some theories for the gender difference, e.g. difference in risk aversion and lack of self-confidence, and points to research which shows that the proportion of women in Norway who thinks they have the necessary competence to become an entrepreneur are significantly lower than among men. She also points out what OECD (2004) is arguing, that increased participation for women into self-employment could make a significant contribution to new business formation, job creation, productivity gains and economic growth. Women who are starting businesses choose a significantly different sector compared to men, and this could be blamed by cultural beliefs about the genders. A rather surprising finding is that having children is no barrier to becoming self-employed for either men or women. Even more surprising, women with children under age of 10 seems to have a higher propensity to become self-employed then women without children. The propensity of becoming self-employed are closely linked to employment status of their spouse or cohabitant. If the partner are out of the labor force, the propensity to become self-employed are higher for both men and women.

2.1.2 Some regulatory aspects on self-employment in Norway

The National Insurance Act (Folketrygdloven (1997)) §1-10 defines self-employment as a person who runs their own business at their own account and risk, rather than working for someone else. Their occupation should also be persistent and have positive net income. When the tax authorities decides whether to permit registration as self-employed or not, they are required to consider the following:

- If the business is of a certain extent.
- If the person has the sole responsibility for the result of the business.
- If the person has employees in his service or is using freelancers.
- If the person runs the business from an established office or building.
- If the person has the financial responsibility for the business
- If the person uses his/her own assets to fund the operation of the business.

In Norway, it is actually not required to have any equity or starting capital to register and start as sole-proprietorship, but as stated above there are still several criteria for the authorities to consider. On the other hand, when starting an incorporation, a minimum amount of capital (30,000 NOK) is required for equity in the new firm. Self-employed persons starting a sole proprietorship are liable for all debt and income and are legally bound to cover any eventual losses from their firm, but this does not apply when starting an incorporation. All self-employed individuals must be enlisted in several governmental registers and be at least 18[3] years of age to start their own incorporation. It is also worth noting that the rules and legislations for social benefits do not cover self-employed persons at the same level as regular employees, see Altinn (2015).

2.2 Benefits of self-employment

Thurik et al. (2008) discusses the relationship between self-employment and unemployment and argues that self-employment can have some positive effects on the level of employment in a society. They find that high rates of self-employment increase entrepreneurial activity and thereby reduce unemployment in subsequent periods, which can be a good thing for the society as a whole. Glocker and Steiner (2007) studied a German pseudo-panel data and found that previous unemployment had positive effects on entry rates into self-employment for both genders, hence governmental legislature supporting an easy and flexible system for the self-employment, may also have positive effects on the level of employment.

Praag and Raknerud (2014) reconsiders the so-called “entrepreneurial puzzle” and compare the transition from self-employment to entrepreneurial activity using matched person-firm data on Norwegian individuals and firms in the period 2002-2011. They find that the average return to entrepreneurship is significantly negative, unless the individual establishes a relatively large incorporation, then the return becomes slightly positive. This is in line with Hamilton (2000), who finds that individuals who switch from wage employment to entrepreneurship gets quite low pecuniary returns

Self-employment can be a driver for innovation and economic growth (OECD, 2004), and most empirical studies measures entrepreneurship in terms of self-employment (Praag and Raknerud, 2014). Romero and Martínez-Román (2012) summarize the key factors of innovation in small businesses to three identifiers: the characteristics of the self-employed individual, the characteristics of the organization and the characteristics of the external environment. They point to personal characteristics that influence innovation in small businesses, such as level of

[3] To start a sole-proprietorship, you have to be at least 15 years old.

education, previous experience and personal motivation. Organization characteristics boils down to technological opportunities within the sector (expected profits from innovations), level of market concentration (number of competitors), cooperation with other firms, number of employees and management style. External environment components related to success and innovation of the self-employed are knowledge spillovers, which allows small businesses to innovate without large investments in research and development (R&D), cultural values in the region and the characteristics of the institutional framework. In addition to the research on determinants of innovation for the self-employed, they also address that 40% of the self-employed workers with small businesses in Spain reported product innovations, and 46% reported process innovations.

2.3 Financing start-ups in Norway

To be able to start a business, it is necessary to have some amount of capital or credit, dependent on how capital intensive the start-up is. Hence, liquidity constraints can ruin the prospects for many individuals eager to start their own business (Disney and Gathergood, 2009). Credit availability and the household's finances is an important issue to undertake, when considering entry into self-employment or stay employed in an existing system with safe barriers. If credit availability is limited, it might get hard to get the business up and running.

There are several ways to deal with the financing of the start-up. Collateral is necessary to be able to get a loan from a commercial bank, and one possibility is to use the house as an asset for this purpose. The surge in house prices in recent years make this a convenient way to get more credit, if your mortgage-to-income ratio and credit worthiness allows it. Accumulated capital or savings from previous years can also make up the equity in the start-up. In recent years, *crowdfunding* has become an increasingly popular way of financing through internet (Belleflamme et al., 2014). Another possibility is venture capital for start-up firms and small businesses with high growth potential. Installment loans for long term needs and overdraft facilities are also possible means for acquiring credit for the shorter term (DNB, 2015). Another possibility is to lease equipment or other necessities, which frees up other working capital. Our focus for this thesis is the acquiring of credit through home-equity refinancing.

If a person has membership in the Norwegian National Insurance Scheme and gets laid off or fired, in general he/she is eligible to unemployment benefits from the Norwegian Labor and Welfare Administration (NAV, 2015). When becoming self-employed, they can be granted these benefits for up to 9 months[4] while developing and establishing your new occupation. There are some initiatives from the government in Norway, but they are mostly directed to innovation and contingent on extensive Research and Development, see The Research Council of Norway (2015). Innovation Norway, a governmental organization for innovation and development of Norwegian businesses also grants subsidies and support, but usually only for innovative start-ups and entrepreneurs (Innovation Norway, 2015).

Several papers has investigated the relationship between household wealth and business entry, and documented a positive relationship. One paper conducting research on an American household survey, suggest that this positive relationship is highly non-linear and only occurring among the top 5th percentile of household income (Hurst and Lusardi, 2004). Hurst and Lusardi also analyzed variation in house price appreciation in different regions and the effect of using home equity as capital for a start-up, but did not find any significant results for the hypothesis that regions with high house price appreciation were more likely to start their own business. Disney and Gathergood (2009) confirms these findings on a UK data set, but their findings also suggest that households with greater wealth are more likely to start up businesses.

2.4 House prices and household Debt

With house prices soaring and significant increase in mortgage debt levels as percentage of household disposable income, there is no doubt the Norwegian housing market has attracted much attention lately. On December 12th 2014, the Norwegian central bank cut the policy rate to 1.25%, after more than two years being at 1.5%. This rate cut was mainly due to a steep decline in oil prices in the second half of 2014 and thereby expectations of lower economic growth in Norway and other oil producing countries, in addition to the low interest rate environment in Scandinavia and the rest of EU. After this, a surge of rate cuts occurred in most of the Norwegian commercial banks, and there are still stiff competition in acquiring as many home mortgage borrowers as possible (NTB, 2015). The debt-to-income ratio for Norwegian households have tripled since the 2000's, fueling up on the ongoing debate whether these high debt levels are sustainable (Lindquist, 2012).

Deregulation in financial markets, falling house prices and high debt levels in the US, is seen as the main reason for the recent worldwide financial crisis in 2007, see Koetter and Poghosyan (2010), Anundsen et al. (2014) and IMF (2009). The crisis brought the relationship between the

[4] Certain other requirements must be met, Altinn (2015) for more information on this.

real economy and financial markets on the agenda around the world. The US household leverage increased sharply preceding the crisis, where the household sector doubled its debt balance in only 5 years. Mian and Sufi (2011) also points out a strong link between house price appreciation and increased household borrowing. Anundsen and Jansen (2013) find a two-way interaction between house prices and household borrowing in the long run. Their research on Norwegian households points out that increased house prices lead to credit expansion, and then credit expansion puts pressure on the house prices. Interest rates also affect house prices indirectly through the credit channel.

Bernanke et al. (1999) introduces an explanation to the housing market fluctuations, a mechanism they call the *financial accelerator.* First, higher house prices increase the amount of credit necessary to buy a house, which in turn strengthens the demand for credit. Second, in general housing loans are secured by the property itself. The net-worth of the households increases as house prices surge, thereby increasing their borrowing capacity. Simultaneously the likelihood of defaults on existing loans reduces, and may give banks motivation to increase their lending. This effect is also prevalent in the Norwegian housing market, as suggested by the cointegration analysis conducted by Anundsen and Jansen (2013).

High levels of debt-to-income has led Lindquist (2012), among others, to ask whether these levels are sustainable in the event of hikes in the loan rate. She evaluates the sustainable household debt in Norway by investigating household’s debt servicing income and their sensitivity to increase in loan rates. She finds that first-time buyers and second steppers groups, which constitutes more than half the household debt in Norway, are vulnerable to rate hikes. Therefore, shocks to income, interest rates or house prices may have serious effects on the financial and price stability, and this has led the Central Bank of Norway to monitor the household closer.

In a fundamental perspective, actual debt should be consistent and not deviate too much from a model prediction based on fundamental explanatory variables and sound economic theory, to ensure the sustainability of the debt. Barnes and Young (2003) defines sustainable debt as “…the level of debt chosen by a household is sustainable whenever the expectations about income growth, house prices, interest rates and other determinants of borrowing that underlie the choice are not falsified or revised.” They used an overlapping generation model on US data to explain different household cohort’s rise in accumulation of debt and assets, and points out that the sustainability of the household debt, crucially depends on the realization of the expectations the households have made their borrowing decision on.

Expectations about higher wages, increasing house prices and lower interest rates may fuel up on the surge in debt levels. The recent years huge profits from loan financed house purchases, due to seemingly ever increasing house prices, may have led the household's willingness for further indebtedness, without equivalent income growth to compensate for the growth in debt (Borgersen and Kivedal, 2012).

2.5 Cash-out / home equity extraction

Mian and Sufi (2011) studies how homeowners borrow in response to increased house prices, in the period 1997 to 2008 in USA. They find significant results of this home equity based borrowing, but the results is not uniform across households. The results suggests that homeowners with high credit card utilization and low initial credit scores have strong tendencies for borrowing against an increase in home equity, and there is weaker tendencies for homeowners with high credit scores. Their results even show that homeowners in the top quartile of the credit score distribution show no tendency for borrowing against increased home equity. A bit surprising finding, as they describe themselves, is that home equity-based borrowing is stronger for younger homeowners. They also look at reasons for equity-based borrowing, and find no evidence for either purchase of new (and better) homes or investment properties. Neither do they find evidence for down payment of expensive credit card debt. By this, they suggest that there is a high marginal private return to borrowed funds.

Su et al. (2014) make use of Norwegian data to research how existing homeowners withdraw cash out of increased home equity by refinancing their mortgage, in response to house price appreciation. They, as we will, call this *cash-out refinancing*. They find that cash-out is present among all groups of households despite the duration of their ownership, income, gender, age, education, household size, number of employed persons in the household, their native origin and the number of kids in the household. Their results suggests that cash-out by existing home owners (in the sample) accounts for at least 36.7% of the total mortgages, which in comparison to mortgages of new home buyers is a substantial part of total household debt. A finding which is consistent with Mian and Sufi (2011) in the American sample is that cash-out is connected to financial difficulties. As reported, the probability of having financial difficulties for the quarter of households with the highest cash-out-to-income ratio is about 45% higher than that for the quarter of households with the lowest cash-out-to-income ratio. In their paper, they compare this cash-out-to-income ratio to mortgage-to-income ratio, where the latter is a widely used measure of household leverage, and suggest that the cash-out-to-income ratio may be a more informative ratio for predicting household financial difficulties. Almaas and Bystrøm (2014) investigates cash-out refinancing in Norway in the period 2001-2012, although they do not find any sufficient evidence for an increasing cash-out over time, they do confirm the presence of cash-out in the Norwegian households.

Benito (2009) studies the decision for home equity withdrawal in the UK over 1993-2003, in the context of a life cycle model of mortgage refinancing (Angelini and Simmons, 2005) and a model where the home-owner maximize present value of utility from holding liquid assets, consumption and mortgage borrowing (Hurst and Stafford, 2004). The key points of such models are that those who borrow against home equity are:

- younger or have a rising income profile, and thus borrow to give a smooth consumption path.
- those who have experienced financial shocks and use housing equity as buffer.
- those who have higher levels of equity in their home, e.g. in response to house price appreciation. This is a precondition for home equity based borrowing.
- those who live in areas with lower local house price volatility.
- those who are liquidity constrained and have few other liquid assets to utilize. This especially when the household is experiencing financial shocks.

In this framework house price appreciation will increase the propensity to withdraw equity at an aggregate level, because of how asset price movements affect borrowing and spending and is related to credit channel models. Another effect on aggregate level is how interest fluctuations will change the benefit of remortgaging, especially an interest rate decrease is associated with home equity extraction.

The empirical findings of Benito (2009) suggests that home equity based refinancing conforms to their choice of economic models. The likelihood of withdrawing equity are high for individuals in their 20s and 30s, with a peak around the age of 40 and a decline in likelihood thereafter. This is what one would expect in the life-cycle frame of reference for consumption path smoothing. The results also show that negative financial shocks have a significantly positive impact on the likelihood of withdrawal, suggesting that in normal times households are less likely to withdraw home equity. Liquidity-constrained households in the beginning of the sample period also show an increased propensity to withdraw equity. Households with more home equity are also more likely to withdraw equity. These are all findings confirming that home equity extraction followed theoretical models within the sample period in the UK. They also show some intuitive findings such as that change in marital status effect the propensity to withdraw home equity.

3 Research Questions

3.1 Motivation

The ongoing debate on the health of the Norwegian debt levels, has led both foreign and domestic researchers and politicians to question the sustainability of the Norwegian housing market, see e.g. Lindquist (2012). Su et al. (2014) suggests that the cash-out effect is much to blame for the recent surge in Norwegian household's debt levels, and with this in mind we think it is necessary with further investigation of these effects. Since there are no sufficient evidence for what cash-out is used for, it is difficult to say much about the broader implications of cash-out contributed debt levels. The present study is a contribution in this respect, and the main purpose of our research is to shed some light on this missing link, by trying to find a relationship between cash-out and spending.

Our idea is simple. We want to find out if cash-out is used as an alternative source of funding for individuals who need capital to start their own small business, which we measure by individuals who transitions from wage employment into self-employment. Even though the question we ask is simple, there is no easy way of measuring such an effect because of the potential self-selection bias of assignment into self-employment. We must overcome the fundamental problem of causal inference[5] which we do with propensity score matching (PSM). This method will enable us to simulate an experimental study, and look for structural differences in the level of cash-out between individuals transitioning into self-employment in comparison with otherwise "equal" individuals.

An idea like this is easily justified, since starting a small business may require a larger amount of capital than normal households have in liquid funds. In Norway many people have their portfolio of investments almost exclusively in their home (Andreassen, 2014), thus after a house price appreciation the value of their home is a natural place to find the capital needed in a start-up phase. Since the self-employed individuals are a small part of the total labor force[6], such an causal effect of self-employment on cash-out would probably be a minor part of the large amount of the total cash-out that Su et al. (2014) finds. Thus, looking for such an effect and quantifying it, may yield surprising results.

[5] See chapter 5.1 for explanation of the fundamental problem of causal inference.
[6] The proportion of self-employees in the workforce is about 7.3%. See chapter 2.1

3.2 An idea in a new direction

As previously discussed, cash-out has first and foremost been associated with financial difficulty of certain home-owners, which could be a financial destabilizer to the national economy. In contrast to such effects, our idea is in a more positive direction. As discussed in chapter 2, there could be several different kinds of benefits to society from self-employed individuals, with examples such as job creation, immigrant's participation in the labor force and in some cases entrepreneurial activity and innovation. It could therefore be argued that if the society's benefits are larger than (or at least not insignificant to) the cost of debt caused by the cash-out of self-employed individuals, dependent on usage, the cash-out in itself may be a cause for good. That said, it is not our purpose to measure the social-economic impacts of self-employment in comparison to debt, and we will leave this as a resource for further debate and research.

3.3 Hypothesis

Our hypothesis for our analysis is whether the decision to become self-employed, which we use to measure start-ups of small businesses, have an effect on the level of cash-out. In the same Norwegian sample as Su et al. (2014) have found evidence for a large cash-out caused debt-to-income ratio, we focus on whether we can find some evidence on what this cash-out is used for. Formally stated our hypothesis is:

Main hypothesis: *The individual's choice of becoming self-employed give rise to a positive and significantly larger amount of "cash-out" when compared to otherwise similar individuals.*

4 Data Material

4.1 Data material

Our analysis is based on data from the Norwegian EU-SILC survey (European Union Statistics on Income and Living Conditions) from 2012. Data were collected mainly through telephone interviews, with some personal interviews as the exception. Additional data from the Norwegian Registration data set is linked to the survey thus completing the data set with information on income and taxes etc. The Norwegian Social Science Data Archives (NSD) makes the data available to researchers and students, after anonymizing the data set. The surveys is carried out annually with a three-year cycle, with different topics each year. The topic in 2012 were on housing, living conditions, and exposure and fear of crime. A representative sample of 11387 people with an age above 16 was pulled from the population. With a response rate of 55.6% leaves a net sample size of 6186. Selection criteria is based on age, gender, education, family size and county (Vrålstad et al., 2013).

We follow Su et al. (2014) and restrict our available data by dropping observations of lesser importance for our study. We first drop all households that do not own their own house, since renters (1,009 observations) and households with shared ownership (728 observations) are irrelevant for our analysis. We further drop observations with missing predicted price (340 observations), missing bought price (190 observations), or missing amount of current mortgage (168 observation). We drop households who have lived in their home for more than 25 years (918 observations) and finally we drop observations of households where no one are employed (312 observations), and are left with a sample of 2,475 observations. For a period over 25 years the household situation may have changed because of exogenous factors such as fast economic growth in Norway, also their possible cash-out may have already been paid back. Furthermore, observations without any employed persons are neither of interest since we only look at difference between those who are employed and those who are self-employed, hence the reason for dropping these variables.

4.2 Descriptive statistics

In table 4.1 we present relevant statistics from the households, including relevant selection criteria variables such as age, gender, education level and more. In table 4.3 the variables listed are defined (more discussion around the variables are given in chapter 6.1). We also list the proportion of households with at least one self-employed individual, and the level of cash-out in the households. The proportion of households with at least one self-employed individual is 10%, note that this is not the real population ratio. Since we have excluded some observations, the reported ratio is the sample-selection ratio[7]. The cash-out variable[8] is winsorized at the 1st and 99th percentile to reduce outlier bias[9].

Table 4.1: Summary statistics for the full sample

Variable	Mean	Median	Min	Max	SD	*N*
Self-employed	0.10	0.00	0.00	1.00	0.30	2,475
Cash-Out	0.53	0.40	-2.34	5.64	1.12	2,190
Age	41.94	42.00	16.00	79.00	12.74	2,475
Gender	0.56	1.00	0.00	1.00	0.50	2,475
Education	1.17	1.00	0.00	2.00	0.82	2,475
Couple	0.82	1.00	0.00	1.00	0.38	2,475
PartnerEducation	0.08	0.00	0.00	1.00	0.27	2,475
NumEmployed	1.75	2.00	0.00	7.00	0.64	2,475
Regions						
Oslo and Akershus	0.24	0.00	0.00	1.00	0.43	2,475
Hedmark and Oppland	0.07	0.00	0.00	1.00	0.25	2,475
Østlandet otherwise	0.19	0.00	0.00	1.00	0.39	2,475
Agder and Rogaland	0.16	0.00	0.00	1.00	0.37	2,475
Vestlandet	0.17	0.00	0.00	1.00	0.37	2,475
Trøndelag	0.10	0.00	0.00	1.00	0.30	2,475
Nord-Norge	0.08	0.00	0.00	1.00	0.28	2,475

Furthermore we present summary statistics in table 4.2, where we divide the statistics into those who are self-employed and those who have reported another type of employment. This gives a

[7] For population ratios of Norway see e.g. Rønsen (2012)
[8] See chapter 5.9 for a description of the cash-out variable.
[9] See appendix C on winsorizing, and figure C.1 for cash-out distribution before and after matching.

brief overview of the main characteristics in the two sub-samples. By visual examination of the data before matching[10] we see that the mean value of cash-out for the self-employed is 0.62 million and for the wage employed it is 0.52 million. This indicates a mean difference between self-employed and wage-employed in respect to cash-out. The table also indicates the age and gender difference, with higher age and more men among the self-employed. There are also a lower mean education level of the self-employed, but there are a higher ratio of self-employed with a partner who have higher education.

Table 4.2: Summary statistics for self-employed and wage-employed

	Self-employed N=243			Wage-employed N=2232		
Variable	Mean	Median	SD	Mean	Median	SD
Cash-Out	0.62	0.39	1.27	0.52	0.41	1.11
Age	47.16	47.00	10.97	41.38	42.00	12.79
Gender	0.67	1.00	0.47	0.54	1.00	0.50
Education	0.79	1.00	0.86	1.21	1.00	0.81
Couple	0.88	1.00	0.32	0.81	1.00	0.39
PartnerEducation	0.23	0.00	0.42	0.06	0.00	0.24
NumEmployed	1.63	2.00	0.68	1.76	2.00	0.63
Regions						
Oslo and Akershus	0.27	0.00	0.44	0.24	0.00	0.42
Hedmark and Oppland	0.08	0.00	0.28	0.07	0.00	0.25
Østlandet otherwise	0.22	0.00	0.41	0.18	0.00	0.39
Agder and Rogaland	0.12	0.00	0.32	0.17	0.00	0.37
Vestlandet	0.14	0.00	0.35	0.17	0.00	0.37
Trøndelag	0.09	0.00	0.29	0.10	0.00	0.30
Nord-Norge	0.08	0.00	0.28	0.08	0.00	0.28

Note: the number of observations of cash-out is 209 for self-employed and 1981 for Wage-employed.

[10] See chapter 5 for description and theory of propensity score matching.

Table 4.3 Definition of variables

Variable	Definition
Self-employed	Equals 1 for self-employed individuals and 0 for wage employed individuals.
Cash-Out	Rough estimate of the level of cash-out refinancing
Age	Age at the time of interview (2012).
Gender	Equals 1 for men and 0 for women.
Education	Equals 0 for primary and lower secondary education, 1 for upper secondary education, and 2 for higher education.
Couple	Equals 1 if individual have a life partner, i.e. married or cohabitant, otherwise 0.
PartnerEducation	Equals 1 if individuals life partner have higher education, otherwise 0.
NumEmployed	The number of employed persons in the household.

5 Theory and empirical approach

We will in this chapter present the theory about the empirical framework of propensity score matching (PSM), and relate it to our application. We will address assumptions of PSM and different matching algorithms used in the analysis. A discussion of data requirements are provided before we conclude the PSM section with theory of sensitivity testing procedures. At last, this chapter ends with a model for cash-out refinancing.

5.1 PSM introduction

Propensity-score matching (PSM) is a widely used approach for estimating casual treatment effects, with applications ranging from evaluation of labor market policies (Heckman et al., 1999), the impact of property taxation on costs control (Borge and Rattsø, 2008) to the effect of water supply on child mortality (Galiani et al., 2005). The method applies to all situations where there is a form of treatment, a group of individuals receiving a treatment and a group not receiving the treatment. The vast amount of applications is due to the diverse nature of possible treatments (Caliendo and Kopeinig, 2008). PSM is well described in the literature, e.g., by Heckman et al. (1997), Heinrich et al. (2010) Becker and Ichino (2002). The seminal references are Rosenbaum and Rubin (1983) and Rosenbaum and Rubin (1985). A recent paper that utilizing PSM method on self-employment is Kaiser and Malchow-Møller (2011), which studies the influence of self-employment on wage.

In evaluation of treatments there are problems due to the fact that we cannot observe both a treated outcome and a non-treated outcome for the same individual at the same time, which is called the *Fundamental Problem of Causal Inference* (Holland, 1986). In the absence of experimental design, taking the mean outcome of non-treated individuals as an approximation for the non-treated outcome, give rise to a possible biased conclusion of causality. This because of how there usually are non-random differences between individuals in the treated and non-treated groups, either by obvious and visual differences or by hidden differences, where this may lead to selection bias (Heinrich et al., 2010). In the case of our application where we measure the level of cash-out of self-employed and wage-employed individuals, the treatment is the choice of becoming self-employed[11]. This choice can lead to a self-selection bias since the treated individuals (those who choose to become self-employed) and the untreated

[11] Treated individuals are those who are self-employed, and untreated are those who are wage-employed.

individuals (those who are wage-employed) will likely differ in personal characteristics. This can be both visual differences like age, gender and education, but also more subtle differences (and possibly unobserved differences) like motivation and risk aversion (Ekelund et al., 2005).

A possible solution to selection bias problems is a matching approach where observable pre-treatment control variables captures relevant differences between any treated and untreated individuals, which can lead to an unbiased estimate of the treatment impact (Dehejia and Wahba, 2002). The basic idea of matching is to create a control group of the non-treated individuals who are similar to the treated individuals with respect to the pre-treatment control-variables (personal characteristics) gathered in a matching vector X. This method increases in difficulty as the size of the X vector grows, and is known as *the curse of dimensionality* where the meaning of the word *similar* becomes less clear, thus the idea of "closeness" in higher dimensions of X is not clearly defined (Heinrich et al., 2010). An approach to reduce the dimensionality problem is the use of balancing scores (Rosenbaum and Rubin, 1983), where one possible type of balancing scores is the propensity score. Propensity score is the probability of being in the treated group given the observed characteristics X, this is the matching method known as propensity-score matching and is our choice of matching method.

5.2 Roy-Rubin model

To formalize the problem of how to measure outcome of an individual dependent on receiving treatment, we introduce the Roy-Rubin-model (Roy, 1951, Rubin, 1974) as presented by Caliendo and Kopeinig (2008) and Heinrich et al. (2010). We denote the treatment effect for an individual i by τ_i, and define this as the difference between the potential outcome in case of treatment and the potential outcome in case absence of treatment:

$$\tau_i = Y_{1i} - Y_{0i} \tag{5.1}$$

Where Y_{1i} is the potential outcome[12] in case of treatment for individual i, and Y_{0i} is the potential outcome in case absence of treatment for individual i. The expected value of τ is known as the Average Treatment Effect (ATE) and gives the mean impact of a program averaged over all the individuals in the population:

$$\tau_{ATE} = E(\tau) = E(Y_1 - Y_0) \tag{5.2}$$

[12] Potential outcome is the level of cash-out in our analysis.

Another and in most cases (as in our study) a more important value is the Average Treatment Effect on the Treated (ATT) (Caliendo and Kopeinig, 2008), this measures the impact of the program on those individuals who are in the treatment group:

$$\tau_{ATT} = E(Y_1 - Y_0 \mid D = 1) \tag{5.3}$$

Where D is a binary variable that denotes the state of the treatment, $D = 1$ for i in treated and $D = 0$ for i in non-treated.

The last value follows naturally and is the Average Treatment Effect on the Untreated (ATU), and measures the impact that the program would have had on those who are in the non-treatment group:

$$\tau_{ATU} = E(Y_1 - Y_0 \mid D = 0) \tag{5.4}$$

The problem with these values (ATT and ATU) are that they are impossible to measure since they rely on counterfactual outcomes. We cannot get a measure of a potential outcome in absence of treatment on an individual who receives treatment. To illustrate this we rewrite 5.3:

$$\tau_{ATT} = E(Y_1 \mid D = 1) - E(Y_0 \mid D = 1) \tag{5.5}$$

Here the second term is unobserved since this is the average outcome that treated individuals would have obtained in the absence of treatment, which is not observed for obvious reasons. An experimental designed study could avoid this problem by utilizing that the potential outcome is independent of treatment status:

$$(Y_1, Y_0) \perp D \tag{5.6}$$

Where we here have the desirable property that characteristics of the individuals are equally distributed between treated and the non-treated groups. Then we have that on average the groups will be identical with the exception of the treatment status. We then get:

$$E(Y_0 \mid D = 1) = E(Y_0 \mid D = 0) \tag{5.7}$$

Here the right-hand side is observable and can be substituted into equation 5.5. In this case running a linear regression on treatment status is valid approach to valuation of the treatment.

Since this property is not present in non-experimental program design, we denote the difference by Δ and rewrite the equation:

$$\Delta = E(Y_1 \mid D = 1) - E(Y_0 \mid D = 0) \tag{5.8}$$

Adding and subtracting $E(Y_0 \mid D = 1)$ yields:

$$\Delta = E(Y_1 \mid D = 1) - E(Y_0 \mid D = 0) + E(Y_0 \mid D = 1) - E(Y_0 \mid D = 1) \tag{5.9}$$

$$\Delta = ATT + E(Y_0 \mid D = 1) - E(Y_0 \mid D = 0) \tag{5.10}$$

$$\Delta = ATT + SB \tag{5.11}$$

Where SB is the selection bias, which is the difference between the counterfactual for treated individuals and the observed outcome for the untreated individuals. The goal of evaluation of ATT is then to reduce selection bias to zero and such get a correct estimation of the parameter[13].

5.3 PSM Assumptions

Conditional Independence Assumption (CIA): Given a set of observable control variables X, which are not affected by treatment, potential outcome are independent of treatment assignment:

$$(Y_1, Y_0) \perp D \mid X \tag{5.12}$$

This implies that conditional on X, treated individuals have the same distribution that non-treated would have experienced if they had participated in the program and vice versa (Heckman et al., 1997). Rosenbaum and Rubin (1983) shows that when potential outcome are independent of treatment conditional on the control variables X, then the potential outcome are also independent of treatment conditional on the balancing score, where the propensity score $P(D = 1 \mid X) = P(X)$, is one possible balancing score. Then the CIA assumption based on propensity score can be written as:

$$(Y_1, Y_0) \perp D \mid P(X) \tag{5.13}$$

Common Support Condition: This requirement rules out the possibility of perfect predictability of D given X:

$$0 < P(D = 1 \mid X) < 1 \tag{5.14}$$

[13] For more detail se e.g. Heckman, Ichimura and Todd (1997), Caliendo and Kopeinig (2008), Heckman, LaLonde and Smith (1999)

This says that treatment outcome given values of X have a positive probability of being both in the treated group and non-treated group (Heckman et al., 1999).

Given that these assumptions holds, the treatment assignment are strongly ignorable (Rosenbaum and Rubin, 1983), and the PSM estimator for the ATT can be written as[14]:

$$\hat{\tau}_{ATT}^{PSM} = E_{P(x)\mid D=1}\{E[Y_1 \mid D = 1, P(X)] - E[Y_0 \mid D = 0, P(X)]\} \tag{5.15}$$

The PSM estimator is the mean difference in outcomes over common support, weighted by the propensity score distribution of the treated individuals (Caliendo and Kopeinig, 2008).

5.4 Model choice

When it comes to choosing a model for estimating the propensity score, there is little advice for choosing a functional form, but there is consensus for choosing a logit or a probit model, see e.g. Rosenbaum and Rubin (1983), Heinrich et al. (2010), Smith (1997), Caliendo and Kopeinig (2008). In principle, there is an opportunity for choosing any binary response model, but the general preferences for logit or probit model rises from the shortcomings of the linear probability model that allow predictions outside the $[0,1]$ bounds of probabilities.

We follow the recommendation of Rosenbaum and Rubin (1983) which seems to be the general case in the literature, and choose to use a logit model for estimating the propensity score. The logit model is in the class of binary response models which takes the form:

$$P(D = 1 \mid X) = F(\beta_0 + \beta X) \tag{5.16}$$

Where the function $F(z)$ $(0 < F(z) < 1)$ have the desired properties which makes it ideal to use as a balancing score. β equals the vector of parameters, X is the vector of control variables and β_0 is the constant term. In the case of a logit model, F is the logistic function with the functional form:

$$F(z) = \frac{\exp(z)}{1 + \exp(z)} \tag{5.17}$$

For all real numbers $z \in \mathbb{R}$ the logistic function is between zero and one (Wooldridge, 2012). Since this is a non-linear model, ordinary least square is not suitable, but usage of maximum

[14] When identifying the ATT it suffice to assume that $Y_0 \perp D \mid P(X)$ and $P(D = 1 \mid X) < 1$ (Smith & Todd 2005)

likelihood estimation is applicable in logit estimation. Further assumptions of logit models are that the observations of the outcome variable are independent of each other, and that we have no strong multicollinarity of the explanatory variables (Tufte, 2000).

The odds of an outcome is measured by $\frac{P(D=1 \mid X)}{1-P(D=1 \mid X)}$, and indicate the ratio of the probability of an outcome of one against the probability of an outcome of zero. By taking the log odds and inserting for $P(D = 1 \mid X) = \frac{\exp(z)}{1+\exp(z)} = p$, we get:

$$ln \frac{p}{1-p} = ln\left(\frac{\frac{\exp(\beta_0 + \beta X)}{1 + \exp(\beta_0 + \beta X)}}{1 - \frac{\exp(\beta_0 + \beta X)}{1 + \exp(\beta_0 + \beta X)}} \right) = \beta_0 + \beta X \tag{5.18}$$

Thus, the coefficients of the logit models approximately estimates the partial effects of a change in a variable on the log odds.

5.5 Choosing matching algorithm

The choice of a matching algorithm in large samples is of lesser importance, because as the sample size increases, all PSM estimators will become closer to comparing exact matches (Smith, 2000). However, when sample size are small the choice of the matching algorithm(s) can be of great importance (Heckman et al., 1997), where there usually are a trade-off between bias and variance (Caliendo and Kopeinig, 2008). The performance difference between each particular matching algorithm depend on the data at hand (Zhao, 2003), and there is no clear rule for determining which algorithm that is more appropriate in each context (Heinrich et al., 2010). We therefore give a short presentation of the most common algorithms[15] and their respective benefits and drawback, to give a more solid foundation for our choice(s).

Nearest Neighbor (NN) Matching: this is the most straight forward matching estimator where individuals from the untreated group is chosen as a matching partner for a treated individual that is closest in terms of propensity score. There are two main categories of NN matching, the first are “matching with replacement” where we allow for an untreated individual to be matched more than once, and “matching without replacement” where an untreated individual only can be matched once. The benefit of replacement is that we get an increased quality of matching

[15] We give short mathematical formulations of matching estimators in appendix F. For further detail of the most common algorithms and mathematical representation of the matching estimators see Smith and Todd (2005)

and reduced bias, but the downside is a higher variance of the estimator due to fewer used observations (Smith and Todd, 2005). Replacement is of particular interest when the propensity score distribution is very different between the treated and untreated (Caliendo and Kopeinig, 2008). Without replacement, we get the opposite results with the benefit of lower estimator variance but a downside of higher bias due to lesser quality of the matches.

Figure 5.1 illustrates the principle of NN matching without replacement. We show two lines of real numbers between zero and one, representing the feasible values of propensity scores for the treated and untreated group. The grey shaded area on the lines represent the distribution of propensity scores for the treated and untreated, and the region where they overlap is the region of common support where we exclude perfect predictability cases. In this example, there are five hypothetical propensity scores for the treated and untreated within the region of common support, represented by the black dots on the lines. The black lines between the dots give the NN matching, and we see that we get some good matches with fairly similar values of propensity score for the treated and untreated. However, on the far right on the line we also get a bad match where the matched untreated propensity score are far away from the treated propensity score.

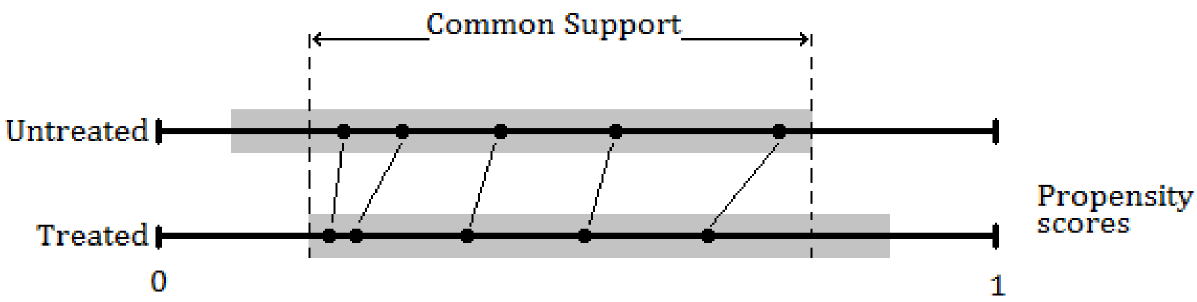

Figure 5.1: Illustration of NN matching with one neighbor without replacement, in the region of common support

Figure 5.2 shows the principle of NN matching with replacement. After the untreated propensity score on the far left have been matched, it is put back in the pool of available propensity scores to match on. Hence, this propensity score get two matches by two different treated propensity scores, and we thus get matches that are closer with respect to the propensity scores and thus a lower bias. However, we also utilize fewer observations, resulting in a higher variance.

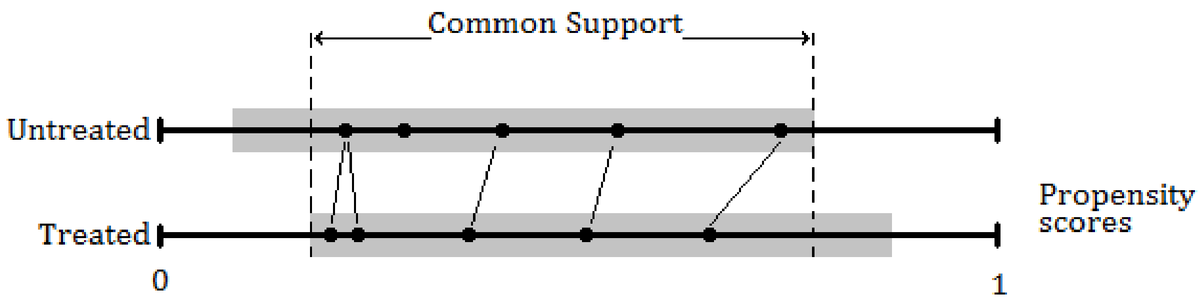

Figure 5.2: Illustration of NN matching with one neighbor and replacement, in the region of common support

The use of oversampling, where more than one nearest neighbor is matched to each treated individual, will reduce variance resulting from using more information to construct the counterfactual for each treated individual, but this will also increase the bias that result from poorer matches on average (Caliendo and Kopeinig, 2008). With oversampling there is a decision to make on how many matches per treated individual we should use and how we should weight them.

Caliper and Radius Matching: to reduce the risk of bad matches that can occur in NN matching when the nearest neighbor is far away, there can be imposed a tolerance level for the maximum propensity score distance (caliper). This have benefits of the same kind as replacements in NN matching since bad matches are avoided, thus a caliper rises the matching quality (Caliendo and Kopeinig, 2008). A drawback of caliper matching is that it is difficult to know a priori what tolerance level should be chosen (Smith and Todd, 2005). Dehejia and Wahba (2002) employ a variant of caliper matching known as “radius matching.” This variant utilizes the mean outcome of all the comparison group members within the caliper, rather than just the nearest neighbor. This method have the same benefits of oversampling in NN matching, but avoids the risk of bad matches, because of how radius matching utilizes all available comparisons inside the caliper, thus making many matches in good cases and fewer matches in bad cases.

We return to our pictorial representation in figure 5.3, where a caliper is restricting the allowed distance between matched propensity scores. The caliper on the far right on the untreated line excludes a bad match (excluded match in dashed line), thus gives less bias but also reduces the number of used observations and gives higher variance.

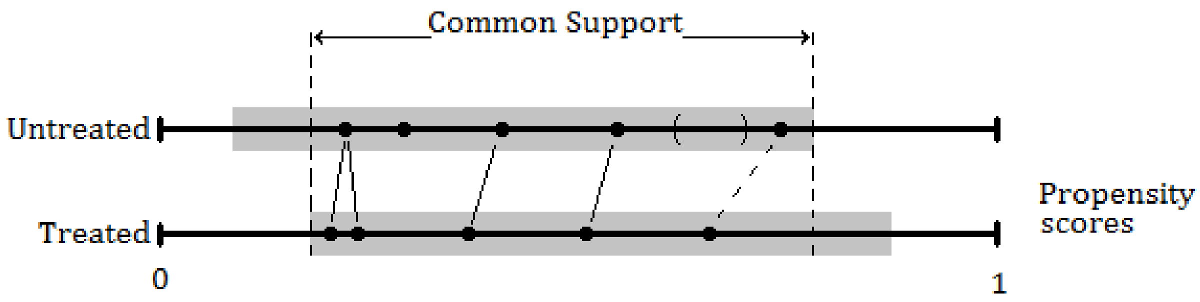

Figure 5.3: Illustration of caliper matching, within the region of common support

Kernel and Local Linear Matching: these matching types differ from those above in that they are non-parametric matching estimators that use weighted averages of all individuals in the control group to construct the counterfactual outcome (Caliendo and Kopeinig, 2008). This has in comparison to the previous methods a major advantage of lower estimator variance due to how these methods use more information. But, a drawback is the possibility of using bad matches. Kernel matching can be seen as weighted regression of the counterfactual outcome on an intercept with weights given by the kernel weights (Smith and Todd, 2005). The weights are determined by the distance between each individual in the control group and the treated individual which the counterfactual is estimated. The estimated intercept provides an estimate of the counterfactual mean. Local linear matching differs from kernel matching in that local linear matching include a linear term in propensity score of treated individuals in addition to the intercept. This advantage of local linear matching is apparent when observations are distributed asymmetrically around the treated observations, or there are gaps in the propensity score. In implementations of these methods one must choose a kernel function and a bandwidth parameter. The choice of kernel function is of lesser importance in practice (DiNardo and Tobias, 2001), but the choice of bandwidth parameter is of greater importance, where there is a tradeoff between high bandwidth with smoother estimated density function, better fit and lower variance between the estimated and the true density function, but higher bandwidth may also smooth away underlying features thus increasing bias (Caliendo and Kopeinig, 2008). Simply put, the choice of bandwidth is a choice between low variance and an unbiased estimate of the true density function[16].

With this short explanation of the algorithms, it becomes clear that we will benefit from reporting the estimates of several matching algorithms with several different algorithmic

[16] A more thorough presentation of kernel and local linear matching can be found in e.g. Smith and Todd (2005)

parameter values. Since asymptotically, different algorithms have the same PSM estimators, we will get a picture of the robustness of our results. If different algorithms provides large leaps between the estimator values, it may indicate that further investigation is necessary to reveal the source of disparity.

5.6 Standard Error estimation

To test for statistical significance of the average treatment effect on the treated (ATT), we have to estimate the standard errors. We use the Stata module PSMATCH2 (Leuven and Sianesi, 2014) which estimates approximate standard error by default on the treatment effects assuming independent observations, fixed weights, homoscedasticity of the outcome variable within the treated and within the control groups and that the variance of the outcome does not depend on the propensity score. See appendix F eq. F.2 for the variance to the ATT estimator.

Another common way of estimating the standard errors is with a bootstrapping approach. Even though bootstrapping is a widely applied method there is little formal evidence to justify the application of bootstrapping (Caliendo and Kopeinig, 2008). Bootstrapping is a random process with resampling of observations, with replacement. Bootstrapping works by re-estimating all the steps for creating the ATT estimator N times. This gives N bootstrapped samples and thus N estimated average treatment effects on treated (ATT). The distribution of these means approximate the sampling distribution of the population mean. We will report both the standard errors as reported by PSMATCH2, and bootstrapped standard errors running 200 simulations per estimation.

5.7 Data and variables: Requirements and drawbacks for PSM

In our analysis we need data that are sufficient for modeling the cash-out refinancing and PSM estimation of ATT for self-employed on cash-out. The data set used (as described earlier) are already proven to be sufficient for modeling of the cash-out effect in Norway, see Su et al. (2014), which makes the choice of this data set ideal in that sense. When it comes to data requirement for propensity score matching, some criteria are not that easy to fulfill.

Implementing matching requires choosing control variables X, such that we satisfy the CIA assumption. Omitting important variables can give a significant increase in the bias of the estimates (Heckman et al., 1997). The choice of control variables should only consider variables that influence participation in treatment, e.g. admission criteria variables, which is key variables that must be fulfilled to participate in the treatment (Heinrich et al., 2010). Other important points to remember is the fact that the variables must either be fixed over time or measured before participation in the treatment, and should stem from the same source (Heckman et al., 1999). Economic theory should back up the chosen variables, but there are also several formal statistical measures to determine a good model specification. The variables that determine the assignment of individuals to self-employed status (variables that influence individual's choice of becoming self-employed) are not all observable (either the known or unknown characteristics), hence we cannot rule out the existence of biased estimators. Known characteristics are more easily implemented since there are many statistical properties for the self-employed in Norway, as discussed in chapter 2.1.1. The unknown characteristics are more complicated, this can be individual qualities such as ability, risk aversion, motivation and intelligence.

Because of the uncertainty of proper model specification, there is a temptation of including all variables that seems reasonable. Hence the question of what is worse, an underspecified model or an overspecified model? Bryson et al. (2002) gives two reasons to be careful with overspecifications. First, including unrelated variables in the model may in fact increase the common support problem. Second, including unrelated variables in the model will not bias the estimators or make them inconsistent, but it may increase variance. These two effects are in fact so strong that matching on a reasonably underspecified model outperforms an over specified model in the estimation of average treatment effects in smaller samples. A convenience with PSM is that poorly estimated propensity scores due to heteroskedastic error terms have little influence on the estimated ATT (Zhao, 2008).

The data source is important when implementing PSM, and the data should either be gathered from the same source or be similarly constructed. In our case a difficult decision is made for us, in the sense that there are not any optimal data source to be found. As described in the last sections, data should either be measured before treatment or be constant over time, which means that the data should be measured when the self-employed individuals decided to become self-employed. In that sense panel data would be beneficial, as panel data with one period before selection into self-employment and one period after selection into self-employment would allow us to relax the CIA assumption and assume that unobserved variables are time-invariant, and thus the effect of these variables can be cancelled out by taking the difference in outcome before and after selection into self-employment. This is the difference-in-difference approach, and is similar to cross-sectional versions, except that outcome is measured in changes (Heinrich et al., 2010).

There are two problems here, first there is not any panel data (to our knowledge) which sufficiently allows us to both do PSM estimation on the self-employed, and model cash-out at the same time. The second problem is that the data which are good enough for cash-out modeling cannot be connected to make panel data because of how the data are anonymized by NSD. We cannot overstate this difficulty enough, since we cannot rule out the presence of unobserved heterogeneity.

5.8 Dealing with unobserved heterogeneity, sensitivity analysis

We test for sensitivity of the ATT by checking what happens if we deviate from the CIA condition, and if the inference about treatment effect may be altered by unobserved factors (Becker and Caliendo, 2007). To give an illustration of how we accomplish this we make a simplifying assumption that the participation probability $P(D_i = 1 \mid X_i, U_i) = F(\beta X_i + \gamma U_i)$ $= P_i$, whereas before, X_i are the observed characteristics of individual i, and U_i are the unobserved characteristics of individual i. β is the effect of the observed X_i, and γ are the unobserved effect of U_i. In experimental program design or in a perfect world without bias, γ will equal zero and the participation probability will be influenced only by the observed control variables. Unfortunately, the world is not perfect and there are bias everywhere, this makes the odds of participation for two individuals with the same value of the control variables possibly different.

We further assume two individuals i and j who have been matched, with F as the logistic function. The odds for the individuals to receive treatment is then $\frac{P_i}{1-P_i}$ and $\frac{P_j}{1-P_j}$ respectively, and then the odds ratio is given by:

$$\frac{\frac{P_i}{1-P_i}}{\frac{P_j}{1-P_j}} = \frac{P_i\left(1-P_j\right)}{P_j(1-P_i)} = \frac{\exp(\beta X_i + \gamma U_i)}{\exp\left(\beta X_j + \gamma U_j\right)}, \qquad where\ 0 \leq U_{i,j} \leq 1 \tag{5.19}$$

We assume that both these individuals have the same value of its observed control variables - as implied by a matching procedure, these terms cancel out and we are left with:

$$\frac{\exp(\beta X_i + \gamma U_i)}{\exp(\beta X_j + \gamma U_j)} = \exp\{\gamma(U_i - U_j)\} \tag{5.20}$$

As an example the unobserved variable $U \in [0, 1]$ could be the degree of risk aversion for an individual, which is especially relevant in the case of self-employment[17] in our analysis. Rosenbaum (2002) shows that the odds ratio of the two individuals receiving treatment is bounded by:

$$\frac{1}{\Gamma} \leq \frac{P_i(1 - P_j)}{P_j(1 - P_i)} \leq \Gamma \tag{5.21}$$

Where $\Gamma = e^{\gamma}$. Both of the matched individuals have the same probability of participating only when the odds ratio, and thus Γ, equals one. If Γ is any larger, then individuals appearing similar with respect to the observed control variables could differ in their odds of participating in the treatment group (becoming self-employed in our application). In fact Γ determines the size of the hidden bias.

To test for such effects in our estimates of ATT, we follow DiPrete and Gangl (2004) who use a Wilcoxon's signed rank test statistic[18] that give upper and lower bound estimates of significance levels at given levels of hidden bias (Γ). This test allows for continuous ATT outcome, which is necessary for estimating cash-out, as opposed to binary outcomes where the Mantel-Haenszel test are more applicable (see e.g. Aakvik (2001)). The test statistic have the form:

$$T = t(Z, r) = \sum_{s=1}^{S} d_s \sum_{i=1}^{2} c_{si} Z_{si} \tag{5.22}$$

In this statistic c_{si} is binary and both $c_{si}(c_{si} \in \{0, 1\})$ and $d_s(d_s \geq 0)$ are functions of r_{si}. The Z_{si} variable registers which of each of the s pairs is treated and equals one if a case is treated and zero otherwise, r_{si} measures the outcome for each case in the S pairs. c_{si} is defined as follows:

[17] See chapter 2.1.1 for discussion of self-employment characteristics.
[18] See Rosenbaum (2002) for exhaustive derivation of this procedure.

$$\begin{aligned} &c_{s1} = 1, c_{s2} = 0 \quad if \quad r_{s1} > r_{s2} \\ &c_{s1} = 0, c_{s2} = 1 \quad if \quad r_{s1} < r_{s2} \\ &c_{s1} = 0, c_{s2} = 0 \quad if \quad r_{s1} = r_{s2} \end{aligned} \tag{5.23}$$

Lastly, d_s is the rank of the absolute difference between r_{s1} and r_{s2}, with average ranks used for ties. When outcome of treated are greater than the outcome of the untreated, the product of c_{si} and Z_{si} cause pairs to be selected. We compare the sum of the ranks of these cases to the distribution of the test statistic under the null hypothesis of no treatment effect. Under the assumption that there is an unobserved effect ($U_{si} > 0$), the test statistic becomes the sum of S independent random variables, where the s^{th} variable equals d_s with probability:

$$\pi_s = \frac{c_{s1}\exp(\gamma U_{s1}) + c_{s2}\exp(\gamma U_{s2})}{\exp(\gamma U_{s1}) + \exp(\gamma U_{s2})} \tag{5.24}$$

And equals 0 with probability $1 - \pi_s$. Though the distribution of $t(Z, r)$ under the null hypothesis are unknown, the distribution are bounded by two known distributions (Rosenbaum, 2002). For any specific Γ the null distribution is upper bounded by T^+ (where T^+ is the distribution when $U_{si} = c_{si}$) and lower bounded by T^- (where T^- is the distribution when $U_{si} = 1 - c_{si}$), which have moments calculated as[19]:

$$\begin{aligned} E\left(T^{+(-)}\right) &= \sum_{s=1}^{S} d_s \pi_s^{+(-)} \\ Var\left(T^{+(-)}\right) &= \sum_{s=1}^{S} d_s^2 \pi_s^{+(-)}\left(1 - \pi_s^{+(-)}\right) \end{aligned} \tag{5.25}$$

Where π_s^+ and π_s^- are the upper and lower bound on π_s given values of U_{si} under T^+ and T^-, and are defined as:

$$\pi_s^+ = \begin{cases} 0 \quad if \quad c_{s1} = c_{s2} = 0 \\ \dfrac{\Gamma}{1+\Gamma} \quad if \quad c_{s1} \neq c_{s2} \end{cases}, \qquad \pi_s^- = \begin{cases} 0 \quad if \quad c_{s1} = c_{s2} = 0 \\ \dfrac{1}{1+\Gamma} \quad if \quad c_{s1} \neq c_{s2} \end{cases} \tag{5.26}$$

[19] Superscript + and – indicates the upper and lower bounds respectively.

Using the constrained values of U_{si} in (5.19), it follows that $\pi_s^- \leq \pi_s \leq \pi^+$ for each $s \in S$. With increasing number of pairs (S), the distributions of T^+ and T^- are approximated with normal distributions. For any specific Γ the bound of the significance level is computed by:

$$z^{+(-)} = \frac{T - E\left(T^{+(-)}\right)}{\sqrt{Var(T^{+(-)})}} \sim N(0,1) \tag{5.27}$$

Where T is the Wilcoxon is signed rank statistic. Increasing Γ decreases the significance, and we thus find the bound where we no longer can conclude that the estimated ATT is not caused by unobserved effects.

5.9 Cash-out modeling

Since we use the same data set as Su et al. (2014), we piggyback on their intuitive model for rough estimation of cash-out, and give a brief presentation here[20]. Their simple definition of cash-out is as follows:

$$CashOut = CurrentMortgage - (OriginalMortgage - PaidMortgage) \tag{5.28}$$

This is intuitive since cash-out is the difference between mortgages an individual have at the time of measurement, minus the original mortgage on the individual's house adjusted for the down payments made since purchase of the home. The data set does not have all these variables available. The only observed variable is the current mortgage, and this makes it necessary to estimate the cash-out. In their model, they assume that the cash-out is a proportion γ_1 ($\gamma_1 \leq 1$) of the house price increase and that the original mortgage is a proportion γ_2 ($\gamma_2 \leq 1$) of the bought price. The paid mortgage equals annual payment (assuming 25-year maturity) multiplied by ownership duration. With these assumptions, we can estimate:

$$\begin{gathered} CurrentMortgage = \gamma_1 \times (CurrentPrice - BoughtPrice) + \\ \gamma_2 \times BoughtPrice\left(1 - \frac{OwnerDuration}{25}\right) \end{gathered} \tag{5.29}$$

Here the first term on the right hand side captures the cash-out, so rewriting this gives:

$$CashOut = CurrentMortgage - \gamma_2 \times BougtPrice\left(1 - \frac{OwnerDuration}{25}\right) \tag{5.30}$$

[20] For further detail of the model as applied to the data set see Su et. al (2014).

The γ_2 parameter varies across mortgages, and we approximate it to be $\gamma_2 = 0.85$. The Home Mortgage Loan Survey conducted by the Financial Supervisory Authority of Norway (2012) reports that 83% of the mortgages had a Loan-to-Value ratio under 85% in 2012 and the average mortgage maturity was 23.1 years. Hence we think that 85% is an appropriate approximation of γ_2 together with a 25 year maturity of home mortgages.

We construct the $CurrentMortgage$ variable by taking $TotalDebt$ which is the sum of all household debt, and scale the number to millions[21]. $OwnerDuration$ is simply 2012 (the year of the survey) minus the reported year of purchase. $BoughtPrice$ is the reported house price. Both $CurrentMortgage$ and $BoughtPrice$ suffer from extreme values, so we have decided to winsorize[22] them at the 1st and 99th percentile to reduce outlier bias.

Table 5.1 list summary statistics of those variables we need to construct the cash-out variable. Where $HaveMortgage$ is a dummy indicating if the household have any mortgage, we see that 88% of the households have mortgage. In Table 5.2 we have divided the statistics giving the mean difference of those with and without mortgage on the variables of interest for cash-out modeling.

Table 5.1 Summary statistics of variables needed to create Cash-Out

Variable	Mean	Median	Min	Max	SD	N
HaveMortgage	0.88	1.00	0.00	1.00	0.36	2,475
BoughtPrice	1.93	1.50	0.00	33.00	1.71	2,475
OwnerDuration	10.10	9.00	0.00	25.00	7.09	2,475

Table 5.2 Summary statistics for Mortgage vs. Non-Mortgage

	Mortgage N=2190			Non-Mortgage N=285		
Variable	Mean	Median	SD	Mean	Median	SD
CurrentMortgage	1.84	1.50	2.60	0.00	0.00	0.00
BoughtPrice	1.96	1.50	1.72	1.68	1.20	1.63
OwnerDuration	9.68	8.00	6.91	13.34	14.00	7.61

[21] Which we will do throughout the text.
[22] See appendix C for a description of winsorizing.

6 Empirical Analysis

6.1 Choice of control variables

The control variables we have chosen for the logit model are based on the characteristics of the self-employed, as described in chapter 2.1.1. It is also necessary to keep in mind the criteria of the propensity score matching procedure, see chapter 5.7. To sum up, we should only include variables that influence participation decision, i.e. self-employment. The choice of variables should also be backed up by economic theory and/or previous findings.

It is important to match on the gender of the self-employed since this is shown in earlier research to be of great importance for the propensity to become self-employed. With regard to PSM criteria the gender variable is of no worry when it comes to measurement before treatment, since it is a very rare phenomenon that individuals change sex. We thus include a dummy variable for the gender of the individual (*Gender*), male as one and female as zero. The age of the self-employed have been shown to be significantly differently distributed than the age of the population of wage employed individuals, this makes age an important matching criteria. One difficulty with the age variable (*Age*) is that we do not know the age of the self-employed when they decided to become self-employed. We only know the age at the measurement date of 2012, and we do not know the length of their employment status. We therefore have to make some assumptions and simplifications for the age variable. We divide age into cohorts of five years, from 20-24, 25-29 and so on, as this decreases the precision of the age variable and thus gives some slack in the matching criteria. The reason for choosing five year intervals for the age variable is consistent with Stambøl (2010) which shows that over 50 percent of the newly started businesses do not survive the first 5 years. We thus make the assumption that the likelihood of individuals to have been self-employed for more than 5 years are small enough to use this as a matching variable. We also include a squared term for the age of the individuals (*AgeSquared*) to capture the change in marginal effects of age.

Further we have chosen the education level (*Education*) as a matching criteria, which is consistent with the statistics on the self-employed, and gives some insight into the individual’s life choice. To meet the PSM requirement of matching of control variables we make the assumption that individuals either are done with education prior to becoming self-employed, or do not undertake any further education while they are self-employed. We believe this assumption is not too far away from reality, and thereby not be a cause for bias.

On the household level, there are several characteristics to match on. If the self-employed have a life partner is important, but again we have to assume that the relationship status have not changed since the individual made the transition into self-employment. This could be a strong assumption, but in regard to how the self-employed on average are older than wage-employed individuals, we make the assumption that the choice of life partner may be more stable for the self-employed than for the rest of the population, or at least have not changed too much since the transition into self-employment. We thus include a dummy variable for people with a life partner (*Couple*) which includes those who are married and those who are cohabitants. We also include a dummy variable that tells whether or not the life partner have higher education (*PartnerEducation*), as this may reflect the social capital, norms and values of the household (Rønsen, 2012). As we saw in chapter 2.1.1 the employment status of other members in the household have an impact on the individuals propensity of becoming self-employed, thus we include a variable for the number of employed people in the household (*NumEmployed*).

We have tested for several effects that did not prove any significance for the individual's choice of becoming self-employed. First we tested for region specific effects, but consistent with Rønsen (2012) these effects are not significant. We further tried to look at the individual's native background, and whether they are immigrants or not (both first- and second-generation immigrants was checked). Neither of those variables are significant which is also consistent with the findings of Rønsen (2012). As a bi-effect of our data selection criteria[23] we are not able to find significant effects of study program on self-employment, which should be present especially within education in the fields of primary industries.

For model specification we use first and foremost a basic textbook econometric approach, as suggested by Heinrich et al. (2010). We start with a bare bones model with only the most important control variables, adding one control variable at the time, keeping only those who are significant and increase the predictive power measured by the pseudo R squared[24]. Table 6.1 reports the results of the models we have chosen based on this approach. The estimators in the full model are all significant, at least at five percent level, and the estimated coefficient signs conforms to the statistical findings of Rønsen (2012). Partners education is significant in our

[23] See chapter 4.1 for description of dropped observations.

[24] Pseudo R squared measure reported by Stata is McFadden's R2. $Pseudo\ \mathrm{R}^2 = 1 - \frac{\mathcal{L}_{model}}{\mathcal{L}_{intercept}}$, where $\mathcal{L}$ is the log likelihood value.

sample, but in the complete sample it may be insignificant as Rønsen (2012) shows. Because of the previously discussed systematic gender difference between the self-employed, we have also listed the results from logit regression on females and males respectively. This shows how the different variables have different estimated implications on the log odds for the different genders. Note that the *Couple* variable and *AgeSquared* is not significant for men and women, respectively. We test for multicollinarity problem in the full sample model with a VIF approach, results in appendix E. Based on this approach we find no issue with multicollinarity in our model.

Table 6.1 Logit model for propensity score estimation

Control Variables	Full sample	Female	Male
Gender	0.4115** (0.1635)		
Age	0.5912*** (0.1714)	0.5424* (0.3097)	0.6757*** (0.2122)
AgeSquared	-0.0412*** (0.0145)	-0.0397 (0.0271)	-0.0472*** (0.0178)
Education	-0.7510*** (0.2628)	-0.6908*** (0.1526)	-0.7815*** (0.1280)
Couple	0.5911** (0.2593)	0.8428** (0.4151)	0.4405 (0.3328)
PartnerEducation	2.1686*** (0.2367)	1.7603*** (0.3540)	2.5083*** (0.3318)
NumEmployed	-0.4786*** (0.1399)	-0.7795*** (0.2342)	-0.3173* (0.1820)
Constant term	-3.3362*** (0.5342)	-2.8228*** (0.8827)	-3.3426*** (0.6590)
N	2030	910	1120
Pseudo R-squared	0.1446	0.1353	0.1504

Standard errors in parenthesis, *** = p <0.01, ** = p < 0.05, * = p<0.1.
The table displays a binary logit estimation for selection into self-employment, for the full sample and the two genders separately.

6.2 Matching quality

There are several different steps for addressing the validity and performance of our matching procedure. We will perform several different balancing tests to ensure that the treatment is independent of unit characteristics after conditioning on observed characteristics. This means that after matching there should be no additional control variable that could be added to the propensity score model that would improve the estimation, and after matching there should be no significant differences between the control variables (Caliendo and Kopeinig, 2008). If we after conditioning on the propensity score finds that there are still some dependence on some control variables, we have to re-specify our model (i.e. move back to square one).

The balancing tests we use are some more formal than others, and least formal is the inspection of the pseudo R-squared. We should see dramatic decrease in the pseudo R-squared after conditioning on the control variables. Since the control variables should indicate no significant difference between the groups after matching, we test for difference in mean of each control variable. We do t-tests on difference in mean before and after matching, and after matching we expect to not reject the null hypothesis of the t-test, which means that we cannot see any significant bias after matching. We will further include a visual representation of standardized bias both before and after matching, as well as we include mean and median bias before and after matching.

The standardized bias is a suitable indicator for assessing the difference in the marginal distributions of the X_i control variables (Rosenbaum and Rubin, 1985). The standardized bias is defined as the difference of sample means for each control variable in treated and untreated group as a percentage of the square root of the average of sample variance in both groups, before and after matching (Caliendo and Kopeinig, 2008). The standardized bias before matching is given by:

$$SB_{before} = \frac{100(\bar{X}_{1i} - \bar{X}_{0i})}{\sqrt{0.5\big(V_1(X_i) + V_0(X_i)\big)}} \tag{6.1}$$

And the standardized bias after matching is given by:

$$SB_{after} = \frac{100(\bar{X}_{1Mi} - \bar{X}_{0Mi})}{\sqrt{0.5\big(V_{1M}(X_i) + V_{0M}(X_i)\big)}} \tag{6.2}$$

Where X_i is the i^{th} control variable and $V(X_i)$ is the variance. Subscript 1 is for treated, 0 is for untreated and M is for the matched sample. An absolute value of standardized bias above 20% should be considered high (Rosenbaum and Rubin, 1985).

Table 6.2 shows the bias reduction from a NN matching procedure with one neighbor and no-replacement. In the unmatched sample, the difference in mean of treated and untreated is significant for all control variables, which means that before matching there is substantial bias between the treated and untreated. However, in the matched sample all differences in means are statistically not significant and the reduced standardized bias are higher than 60 percent for all variables. After matching, the absolute value of the standardized bias is below 20% for all control variables.

Table 6.3 shows pseudo R-squared, mean bias and median bias before and after matching. Before matching the pseudo R-squared are 0.1446 and significant at a 1 percent level. In a re-estimation of the propensity scores on the matched sample, the pseudo R-squared is equal to 0.008 and no longer significant, indicating that the variables do not have any explanation power after matching. For the model as a whole in the matched sample, the mean bias is reduced from 33.6 to 7.7. The table also includes Rubin's R and B which shows ratio of treated variance over untreated variance of propensity score, and the number of standard deviations between the means of the groups respectively (Rubin, 2001). The R ratio should be between 0.5 and 2 and the B should be below 25%, and both of these condition are met for the matched sample.

Table 6.2 Bias reduction in the control variables

Variable	Unmatched Matched	Mean Treated	Mean Untreated	%Bias	%Reduced Bias	t-test t	t-test P > |t|
Age	U	5.7703	5.1697	28.4		3.88	0.000
	M	5.7644	5.9615	-9.3	67.2	-0.95	0.344
AgeSquared	U	37.722	31.233	26.6		3.77	0.000
	M	37.668	40.048	-9.7	63.3	-0.96	0.340
Gender	U	0.6363	0.5420	19.2		2.60	0.009
	M	0.6394	0.6298	2.0	89.8	0.20	0.839
Education	U	0.7799	1.3185	-67.1		-9.66	0.000
	M	0.7836	0.8846	-12.6	81.3	-1.23	0.271
Couple	U	0.8899	0.8286	17.7		2.26	0.024
	M	0.8894	0.9086	-5.5	68.6	-0.65	0.516
PartnerEducation	U	0.2248	0.0335	59.4		12.08	0.000
	M	0.2211	0.1778	13.4	77.4	1.10	0.271
NumEmployed	U	1.6746	1.7814	-16.8		-2.37	0.018
	M	1.6827	1.6731	1.5	91.0	0.15	0.878

The table presents mean values of treated and untreated before and after matching for each control variable. %Bias is the standardized bias, and %Reduced Bias is the reduction in standardized bias after matching. t values are reported from test for difference in means of treated and untreated.

Table 6.3 Difference between unmatched and matched

Sample	Pseudo R Squared	Likelihood Ratio chi2	P > chi2	Mean Bias	Median Bias	Rubin's B	Rubin's R
Unmatched	0.1446	191.10	0.000	33.6	26.6	96.5	2.29
Matched	0.0080	4.43	0.729	7.7	9.3	20.6*	1.45*

The table presents the difference of Pseudo R squared, mean bias, median bias and Rubin's R and B for unmatched and matched sample.
* if B < 25%, R in [0.5, 2]

Figure 6.1 gives a visual illustration of the standardized bias before and after matching, for NN matching without replacement. We see how the standardized bias is further away from zero before matching, than after. This shows visually how effective the matching is, and we can conclude that the propensity score acts as a balancing score. The balancing properties are fulfilled across all of our matching algorithms, and show similar balancing results as in the NN matching provided in this section.

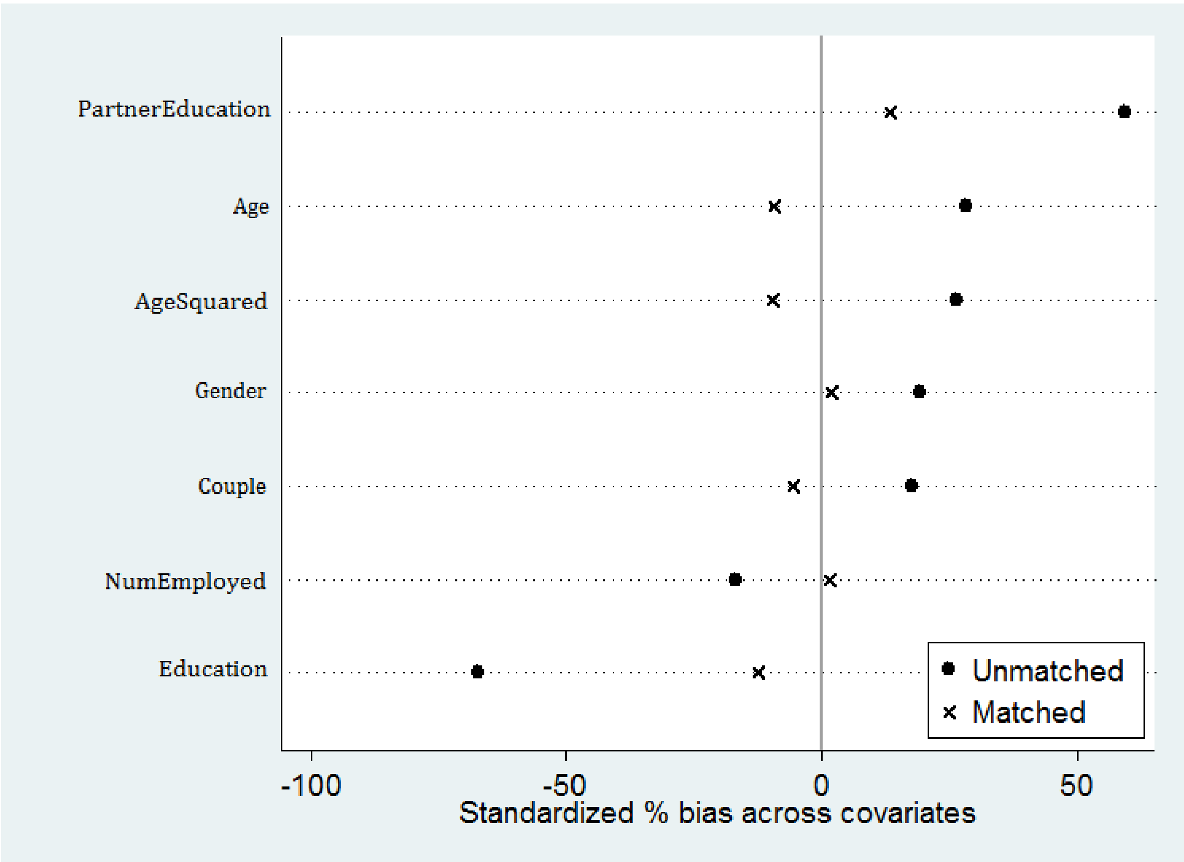

Figure 6.1: Pictorial illustration of difference in Standardized bias, before and after matching

6.3 Common Support

The average treatment effect is only valid over the region of common support. To check that the common support condition holds we first and foremost check the distributions of treated and non-treated groups before and after matching. This is the most straight forward way of checking for common support, and as Lechner (2008) reasons, it is so easy to spot a common support problem that complicated formal test-statistics are of lesser importance. We have also ensured that the common support condition is met by setting a min/max criteria, as this excludes observations that lies outside the region of common support.

Figure 6.2 shows the distribution of the treated and the untreated before and after matching. In the top panel (before matching), we see a clear deviation between the distributions, which in the lower panel (after matching) is much more similar. There is also an overlap between the distributions i.e. there are no clear difference in the minima and/or maxima of the distributions. By visual inspection, we can see that the overlap condition is met, for NN matching without replacement.

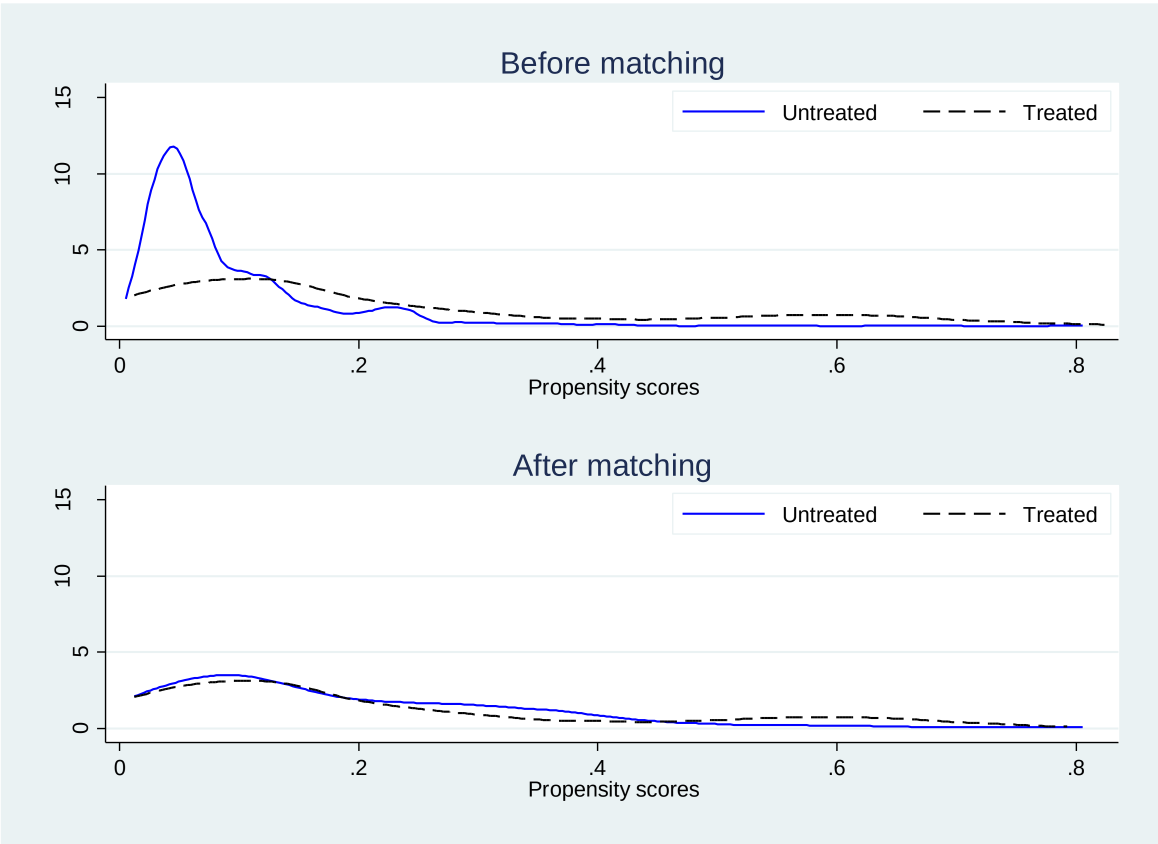

Figure 6.2: Propensity score distribution of Self-employed and Wage-employed, before and after matching

Another aspect we can read out of the distribution (before treatment), is that the density of the propensity scores (for treated and untreated) have large thin tails. We will address this further in the next section.

6.4 Average Treatment effect on Treated

In table 6.4 we present estimated average treatment effect on treated (ATT) for NN matching and caliper matching, and table 6.5 gives radius matching, kernel and local linear matching. The estimated ATT's are interpreted as the level of cash-out (in millions) caused by the self-selection into self-employment. As discussed in chapter 5.5, we see the benefit and drawback of each matching approach. Those algorithms utilizing fewer matches, i.e. less oversampling, have less bias in the estimator, but high variance. Those algorithms utilizing more oversampling have a bias in estimators, but lower variance. There is significance for some of the algorithms, but there are too much disparity between them to conclude that our estimations are robust. This means that we cannot immediately say with confidence that the choice of becoming self-employed is a cause for cash-out, but we will investigate this further before making any conclusions.

The NN-matching without replacement (in table 6.4) show a significant ATT of 0.2350 at 5 percent level. When we allow for replacement, we get better matches, as some control observations are used on multiple observations in the treatment group. This reduces the bias in the ATT estimator, but with a tradeoff of higher variance as fewer observations in the control group are used. The NN matching with one neighbor and replacement have a coefficient of 0.3257, which is significant at the 5 percent level. When we use NN with oversampling (i.e. more matches per treated) we get increased bias, since there is a larger distance in propensity scores between the matched pairs. The benefit of oversampling is reduced variance, but in this case not enough to compensate for the increased downward bias in the estimator.

There is a better tendency within the caliper-matching algorithm. Since the allowed distance between the propensity scores of the matches are set with a fixed caliper, we do not get as many bad matches, thus the estimator has less bias. This tendency increases with more constraining calipers, i.e. a smaller caliper. As we restrict the caliper, the ATT estimators become more significant and robust to oversampling, in contrast to NN-matching without a caliper.

Table 6.4 Estimated ATT for NN and Caliper matching

	Nearest Neighbors			
Algorithm	1	2	5	10
NN Matching without replacement	0.2350 (0.1107)** [0.1270]*			
Observations Treated/Untreated	208/208			
NN Matching With replacement	0.3257 (0.1537)** [0.1716]*	0.2502 (0.1282)* [0.1435]*	0.1699 (0 .1192) [0.1196]	0.1499 (0.1109) [0.1211]
Observations Treated/Untreated	208/96	208/183	208/412	208/695
Caliper Matching (0.01)	0.3582 (0.1545)** [0.2010]*	0.2411 (0.1343)* [0.1417]*	0.2049 (0.1303) [0.1193]*	0.1590 (0.1252) [0.1119]
Observations Treated/Untreated	196/95	196/180	196/406	196/677
Caliper Matching (0.001)	0.3771 (0 .1581)** [0 .1910]**	0.2758 (0 .1296)** [0 .1568]*	0.2460 (0 .1221)** [0 .1441]*	0.1882 (0 .1170) [0 .1349]
Observations Treated/Untreated	170/89	170/165	170/360	170/587
Caliper Matching (0.0001)	0.3826 (0 .1670)** [0 .2082]*	0.2820 (0 .1357)** [0.1806]	0.2672 (0.1294)** [0 .1296]**	0.2164 (0 .1233)* [0 .1216]*
Observations Treated/Untreated	161/81	161/149	161/304	161/471

Standard error in parenthesis, bootstrapped standard error in square brackets.
*** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$

When we use all available matches within a caliper we get radius matching (table 6.5). In our case, this increases the bias of the estimators, and we no longer get significant results. The bias decreases with more restrictive radiuses, but not enough to offset the relatively large variance. As mentioned in chapter 6.3, there is a long right tail and high density in the left part of the propensity score distribution for the untreated in the unmatched sample. This makes the propensity score distribution for the untreated sensitive for oversampling in the left part of the distribution. Thus, when we utilize large oversampling, e.g. radius matching, the matching procedure does not equalize the distribution of propensity scores for treated and untreated. This implies that the propensity distribution is not robust to heavy oversampling. Neither radius, kernel nor local linear produces any significant results. The fact that local linear matching give less significant estimators than kernel matching should raise some worry. We argue that kernel (and radius) matching give bad matching because of the large difference in the distribution, but local linear should, and do, compensate for the shortcomings of kernel matching, with its use of a linear term in addition to the constant term[25]. This raises the need for further investigation of the propensity score distributions.

Table 6.5 Estimated ATT for Radius, Kernel and Local linear matching

	Radius			
Algorithm	0.01	0.005	0.001	0.0005
Radius Matching	0.1186 (0 .1241) [0.1174]	0.0771 (0.1162) [0 .1201]	0.1658 (0.1150) [0.1252]	0 .1866 (0 .1172) [0 .1273]
Observations Treated/Untreated	196/1806	184/1788	170/1375	168/1273
	Bandwidth			
Algorithm	0.1	0.2	0.3	0.4
Kernel matching	0.1602 (0.1106) [0.1021]	0.1501 (0.1039) [0.0980]	0.1328 (0. 0978) [0 .0925]	0.1238 (0.0942) [0.0962]
Observations Treated/Untreated	208/1821	208/1821	208/1821	208/1821
Local Linear	0.1511 (0.1537) [0.1165]	0.1264 (0.1537) [0.1279]	0.1033 (0.1537) [0.1059]	0.0881 (0.1537) [0.1055]
Observations Treated/Untreated	208/1821	208/1821	208/1821	208/1821

Standard error in parenthesis, bootstrapped standard error in square brackets.
*** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$
For kernel and local linear matching, Epanechnikov kernel function are used.

In table 6.6 we show the distribution of propensity scores before and after matching, with NN-matching without replacement and radius matching. Here we see that the unmatched distribution of propensity scores have significantly different properties between treated and untreated. In the untreated group, we see a large kurtosis, which reflects the few observations in the upper percentiles, while the treated have a low kurtosis and in fact almost no excess

[25] See chapter 5.5 on matching algorithms.

kurtosis. By matching we want to equalize the distribution between treated and untreated, and we see that in the NN-matching method we get a smoother distribution difference, although not perfect, it is much smoother than the distribution difference in the radius matching method.

Table 6.6 Distribution of propensity scores for untreated and treated

	Unmatched		Matched (NN)		Matched (Radius)	
percentiles	Untreated	Treated	Untreated	Treated	Untreated	Treated
min	0.0049	0.0122	0.0122	0.0122	0.0122	0.0122
p(1)	0.0125	0.0218	0.0218	0.0218	0.0218	0.0218
p(5)	0.0218	0.0363	0.0363	0.0363	0.0292	0.0363
p(10)	0.0264	0.0436	0.0436	0.0436	0.0293	0.0426
p(25)	0.0392	0.0666	0.0666	0.0666	0.0426	0.0612
p(50)	0.0612	0.1447	0.1447	0.1447	0.0612	0.1075
P(75)	0.0112	0.3212	0.2888	0.3212	0.1075	0.2034
P(90)	0.1963	0.5697	0.3859	0.5697	0.1700	0.3212
P(95)	0.2366	0.6418	0.4964	0.6418	0.2267	0.5428
P(99)	0.4089	0.7327	0.6907	0.7327	0.3212	0.6907
Max	0.8054	0.8239	0.8054	0.7924	0.7924	0.7924
mean	0.0887	0.2271	0.1907	0.2242	0.0843	0.1512
SD	0.0818	0.2063	0.1533	0.2026	0.0678	0.1392
Skewness	2.9519	1.1494	1.3934	1.1387	3.3325	2.3292
Kurtosis	16.9755	3.0728	5.1131	3.0192	24.22	9.0399
N	1821	209	208	208	1273	168

Distribution of propensity scores for the treated and untreated before and after matching. The matched distributions are NN matching without replacement, and radius matching with radius of 0.0005.

6.4.1 Trimming the propensity score distribution

The keen observer notice that the propensity score distribution of the untreated in the unmatched sample in table 6.6, smooth out below the 95th percentile, reflecting the few observations with large distance between the propensity scores. Smith and Todd (2005) suggest setting a trimming level to ensure that the density are strictly positive and exceeds zero by a threshold determined by the trimming level. This will restrict the region of common support by dropping propensity scores within regions of positive but very low densities. We set a trimming level and trim away approximately the top 5 percentiles of the untreated propensity scores, and the top 40 percent of the treated are further trimmed away by common support condition. The new maximum propensity score after trimming is 0.2318, and we see better balancing properties than within the full sample, see appendix D for balancing test tables.

Table 6.7 and 6.8 shows the result of re-estimation of ATT in trimmed sample. In these two tables we see an increased robustness across the different matching algorithms. The estimated ATT ranges between 0.2 and 0.5 indicating a substantial amount of cash-out contributed by the self-selection into self-employment. In chapter 7 we will give a discussion of an interesting possible reason for the increased robustness in the lower part of the propensity score distribution. Because of the previously discussed limitations in our analysis, we have to be careful not to over interpret our results, and in the next section we will test for unobserved heterogeneity in our model specification.

Table 6.7 Estimated ATT for NN and Caliper matching after trimming

	Nearest Neighbors			
Algorithm	1	2	5	10
NN Matching without replacement	0.4258 (0.1457)*** [0.1703]**			
Observations Treated/Untreated	126/126			
NN Matching With replacement	0.5506 (0.1672)*** [0.2313]**	0.3748 (0.1549)** [0.1804]**	0.3228 (0.1427)** [0.1566]**	0.2995 (0.1327)** [0.1463]*
Observations Treated/Untreated	126/66	126/131	126/315	126/559
Caliper Matching (0.01)	0.5506 (0.1672)*** [0.2177]**	0.3748 (0.1549)** [0.1889]**	0.3228 (0.1427)** [0.1690]*	0.2414 (0.1350)* [0.1532]
Observations Treated/Untreated	126/66	126/131	126/315	126/559
Caliper Matching (0.001)	0.5506 (0.1672)*** [0.2338]**	0.3875 (0.1480)*** [0.1841]**	0.3265 (0.1429)** [0.1598]**	0.2462 (0.1367)* [0.1468]*
Observations Treated/Untreated	126/66	126/129	126/300	126/505
Caliper Matching (0.0001)	0.5448 (0.1749)*** [0.2320]**	0.3744 (0.1539)** [0.1900]**	0.3380 (0.1502)** [0.1738]*	0.2637 (0.1429)* [0.1481]*
Observations Treated/Untreated	121/61	121/117	121/250	121/399

Standard error in parenthesis, bootstrapped standard error in square brackets.
*** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$

Table 6.8 Estimated ATT for Radius, Kernel and Local linear matching after trimming

	Radius			
Algorithm	0.01	0.005	0.001	0.0005
Radius Matching	0.2136 (0.1282)* [0.1306]*	0.2062 (0.1290) [0.1310]	0.2381 (0.1332)* [0.1462]	0.2534 (0.1350)* [0.1609]
Observations Treated/Untreated	126/1677	126/1666	126/1266	126/1169
	Bandwidth			
Algorithm	0.1	0.2	0.3	0.4
Kernel matching	0.2674 (0.1272)** [0.1296]**	0.2783 (0.1272)** [0.1288]**	0.2821 (0.1271)** [0.1366]**	0.2826 (0.1271)** [0.1293]**
Observations Treated/Untreated	126/1787	126/1808	126/1812	126/1816
Local Linear	0.2783 (0.1672)* [0.1306]**	0.2799 (0.1672)* [0.1301]**	0.2820 (0.1672)* [0 .1339]**	0.2823 (0.1672)* [0.1408]**
Observations Treated/Untreated	126/1787	126/1808	126/1812	126/1816

Standard error in parenthesis, bootstrapped standard error in square brackets.
*** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$
For kernel and local linear matching, Epanechnikov kernel function are used.

6.5 Sensitivity

To test for sensitivity against unobserved heterogeneity we use the Rosenbaum bounds method as explained in chapter 5.8. Table 6.9 gives the results of such a test on NN matching with replacement and one neighbor, and a caliper matching method with one neighbor. In the calculation of the test statistic we use the Stata module RBOUNDS which requires the use of matched 1x1 pairs only (Gangl, 2004). The test in table 6.9 is for the full propensity score distribution, with tested bounds (Γ) from 1 up to 1.35 in steps of 0.05. We are interested in sensitivity for overestimation of average treatment effect on treated, but have included a test for underestimation for the sake of completion.

The least robust estimation of ATT to the presence of unobserved bias, is the NN(1). This test shows that the upper bound of 1 (no hidden bias) on the odds ratio of two matched individuals are significant and the critical level at which we would have to question the estimated positive effect is between 1.10 and 1.15. This means that an unobserved control variable have to alter the odds ratio of treatment assignment to differ by a factor of about 1.15 between the treated and the untreated, in order to undermine the estimated ATT. By utilizing the caliper matching method, the Γ needs to be higher than 1.20 to disrupt our results in the matching analysis. This indicates that robustness of the ATT in respect to hidden bias is higher for caliper matching, because of how the caliper matching restricts the amount of bad matches. The lower bound at 1 is also significant, and with higher Γ the significance level increases, which is expected since we have estimated a positive treatment effect.

The test indicates that we have some sensitivity to hidden bias since the bounds are fairly low, and thus we may have overestimated the treatment effect. The sensitivity to hidden bias comes as no surprise as we have discussed the limitations of the available data, but we still think we have indicated a possible causal connection between self-selection into self-employment and cash-out refinancing that would be of interest for further investigation.

Table 6.9 Sensitivity test of unobserved heterogeneity

	NN(1)		NN(1), Caliper (0.0001)	
Γ	p^+	p^-	p^+	p^-
1.00	0.0062	0.0062	0.0030	0.0030
1.05	0.0140	0.0025	0.0066	0.0013
1.10	0.0282	0.0009	0.0131	0.0005
1.15	0.0513	0.0003	0.0239	0.0005
1.20	0.0854	0.0000	0.0401	0.0002
1.25	0.1318	0.0000	0.0631	0.0000
1.30	0.1903	0.0000	0.0938	0.0000
1.35	0.2594	0.0000	0.1323	0.0000

Γ : odds of differential assignment due to unobserved factors
p^+ : upper bound significance level
p^- : lower bound significance level

Table 6.10 gives the same sensitivity analysis as above, but now on the ATT from the trimmed propensity score distribution. In comparison to the untrimmed propensity score distribution, we now see a higher significant bound. For the NN(1) algorithm the bound is now between 1.45 and 1.5, and for the caliper algorithm the bound is between 1.4 and 1.45. This shows that when we have trimmed away the lower densities of the propensity score distribution, we have a more robust result in respect to unobserved heterogeneity.

Table 6.10 Sensitivity test of unobserved heterogeneity after trimming

	NN(1)		NN(1), Caliper (0.0001)	
Γ	p^+	p^-	p^+	p^-
1.00	0.0003	0.0003	0.0004	0.0004
1.05	0.0007	0.0001	0.0009	0.0001
1.10	0.0015	0.0000	0.0020	0.0000
1.15	0.0029	0.0000	0.0038	0.0000
1.20	0.0053	0.0000	0.0068	0.0000
1.25	0.0091	0.0000	0.0113	0.0000
1.30	0.0148	0.0000	0.0179	0.0000
1.35	0.0228	0.0000	0.0271	0.0000
1.40	0.0336	0.0000	0.0393	0.0000
1.45	0.0478	0.0000	0.0549	0.0000
1.50	0.0656	0.0000	0.0742	0.0000

Γ : odds of differential assignment due to unobserved factors
p^+ : upper bound significance level
p^- : lower bound significance level

6.6 The size of cash-out caused by the self-employed

Su et al. (2014) roughly quantifies the size of the cash-out[26] to be 36% of the current mortgage, which is a significant part of the total current mortgage of the Norwegian households. We have studied the cash-out behavior for those who become self-employed, and we find that their total cash-out within our sample sum up to about 163 million. By relating this to the total cash-out in our sample of about 1509 million[27], the cash out contributed by the self-employed constitutes about 11% which is a substantial amount contributed by a small part of the labor force. The total current mortgage of the Norwegian households in our sample is about 4025 million which in fact makes the cash-out of the self-employed about 4% of the total current mortgage.

When we look at the ATT estimates of cash-out, we find to what extent the cash-out of the self-employed is caused by the self-selection into self-employment. Even though the estimated ATT differs for the different algorithms, they seems to circle around 0.3, and since cash-out in itself is a rough estimate the ATT will be rough estimates too. By taking size of the ATT in comparison to the total cash-out we find that the ATT on cash-out constitutes about 5% of total cash-out. This is lower than total percentage of cash-out from the self-employed, and reflects the amount of cash-out that are used to fund the start-up of businesses, thus the causal effect on cash-out behavior of becoming self-employed.

[26] Setting negative cash-out estimates to zero when computing the sum of the lower bounds.
[27] Note that these levels are a little lower than what found by Su et al. (2014) since we have restricted our sample further.

7 Further Discussions

7.1 Gender Difference

In the last chapter we found that the average treatment effect on treated were much more significant and robust across the different matching algorithms in the lower part of the propensity score distribution (when we trimmed the propensity scores). We have made a calculated guess that this could be caused by the gender difference of the self-employed as previously discussed, where women have a lower propensity to become self-employed than men. Hence, we want to further see if there is some difference in the average treatment effect on treated between men and women in an informal analysis. The estimated propensity scores come from the logit model for each gender, see table 6.1 in chapter 6.1.

In appendix A we have given summary statistics for the self-employed and each gender separately (table A.1). We find that the mean cash-out of men who are self-employed (0.51 million) are at a similar level as the mean cash-out of the wage-employed (0.52 million, from table 4.2 in chapter 4.2). However, women have a much larger mean cash-out in comparison, at a level as high as 0.81 million. There are some mean age difference, but the only visually considerable mean difference between the genders of the self-employed are the level of cash-out.

In appendix B we have given ATT tables for women and men separately. In these tables there is a striking difference between the genders. For the male self-employed individual we cannot see any significant ATT (table B.1 and B.2). This means that in our sample, and with our matching criteria, the male participants do not have a causal connection between self-employment and cash-out refinancing. However, the female participants seems to have much more significant and positive ATT values, and more robust results over the different algorithms (table B.3 and B.4). This means that the act of becoming self-employed for women seems to be a cause of cash-out refinancing. This gives an indication of different behavior or opportunities for the female self-employed. As noted by Rønsen (2012), an explanation could be that female entrepreneurs attract less capital in the start-up phase[28], thus maybe home equity could be a more inviting solution to raise start-up capital for women than for men. Another possible explanation could be attributed to psychological factors like risk aversion (see e.g. Ekelund et al. (2005), Brown et al. (2011)) and how women in general are more risk averse than men. One could justify such claims by how home-equity may feel like a safer place to look for funding

[28] See background literature, chapter 2.1.1

than more conventional sources. At this point the gender difference in ATT are mere calculated speculation, and first and foremost a resource for further investigation as this can have some implications for policy decisions.

7.2 Policy implications

Su et al. (2014) discusses policy implications in regard to the high cash-out contributed debt levels, and suggests that the cash-out to income ratio may be more informative than the mortgage to income ratio, hence housing policies should pay more attention to regulating home equity based refinancing. With the discussion in chapter 7.1 and chapter 6.6 in mind, such regulations may have adverse effects on women's opportunity to become self-employed. Rønsen (2012) suggest that a higher women to men ratio among the self-employed may be beneficial to economic growth and job creation, thus a policy regulation should not limit the available liquidity for women, who more than men seems to rely on cash-out as a financing option of start-ups. A solution could be complimentary governmental support to female entrepreneurs, which may decrease the amount of cash-out from self-employed women.

8 Conclusion

In our analysis, we have raised an interesting question about a causal relationship between the self-selection into self-employment and the level of cash-out refinancing. The sample data show that self-employed individuals have about 0.1 million larger mean cash-out than the wage-employed. Using propensity score matching we address the issues of self-selection bias caused by non-random differences between the self-employed and the wage-employed. The matching results indicates that the average amount of cash-out caused by self-selection into self-employment ("average treatment effect on the treated") differs between 0.1 and 0.3 million among the different matching algorithms. The significance level of the estimates is sensitive to different matching algorithms, which indicates that the results are not robust. We have overcome data limitations with some empirically sound assumptions regarding control variables. However, this leaves room for some bias caused by unobserved heterogeneity, which we have tested for with a sensitivity analysis, indicating that we should use some caution when interpreting the estimated average treatment effect.

Next, we investigate a trimmed sample where we cut away the lower densities of the propensity score distribution. Within this trimmed sample we get robust results with estimated average treatment effect on the treated ranging from 0.2 to 0.5 million in cash-out. An advantage with this result is a decreased sensitivity for unobserved heterogeneity compared to the untrimmed sample. Further, we informally hypothesize that the increased average treatment effect in the trimmed sample is caused by a gender difference. We investigate the mean cash-out of men and women in the full sample, which shows that women have a substantial larger mean cash-out than men. By matching in the subsample of men and women, we find no significant average treatment effect on treated for men, but for women there are significant and robust average treatment effect on treated between 0.4 and 0.6 million.

We conclude that our data indicate a causal relationship between selection into self-employment and cash-out, especially for women. The size of the cash-out caused by the self-selection into self-employment are roughly estimated to be about 5% of the total cash-out, which is a sizeable amount of the total cash-out. Thus, our thesis have explained some of the ambiguity around the cash-out behavior, in that we have found a causal relationship quantified as a measureable amount of total cash-out. But, due to our data limitations we urge for more research on this causal relationship to explain some of the effects from cash-out, which Su et al. (2014) links to the recent surge in the Norwegian household's debt levels.

9 References

Aakvik, A. (2001): "Bounding a matching estimator: the case of a Norwegian training program". *Oxford bulletin of economics and statistics,* 63**,** 115-143.

Almaas, S. S. & L. S. Bystrøm. (2014): "The Cash-out Refinancing in the Norwegian Housing Market". MSc, Norwegian University of Science and Technology.

Altinn. (2015): "Dagpenger under arbeid". Available: https://www.altinn.no/en/Start-and-Run-a-Business [Accessed 13 March 2015].

Andreassen, H. M. (2014): "Norske husholdninger: Aksjer eller eiendom?". Available: http://www.vff.no/Internett/Statistikk_og_fakta/Presentasjoner/filestore/HMAndreassen_HusholdningeneEiendomelleraksjer.pdf [Accessed 8th May 2015].

Angelini, V. & P. Simmons (2005): "Housing debt, employment risk and consumption". *University of York Discussion Papers in Economics*, no. 2005/07.

Anundsen, A. K., K. Gerdrup, F. Hansen & K. Kragh-Sørensen (2014): "Bubbles and crisis: The role of house prices and credit". *Norges Bank,* working paper 14/2014.

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

Barnes, S. & G. Young (2003): "The rise in US household debt: assessing its causes and sustainability". Bank Of England. report no. 206.

Becker, S. O. & M. Caliendo (2007): "Mhbounds-sensitivity analysis for average treatment effects". *The Stata Journal,* 7**,** 71-83.

Becker, S. O. & A. Ichino (2002): "Estimation of average treatment effects based on propensity scores". *The Stata Journal,* 2**,** 358-377.

Belleflamme, P., T. Lambert & A. Schwienbacher (2014): "Crowdfunding: Tapping the right crowd". *Journal of Business Venturing,* 29**,** 585-609.

Benito, A. (2009): "Who Withdraws Housing Equity and Why?". *Economica,* 76**,** 51-70.

Bernanke, B. S., M. Gertler & S. Gilchrist (1999): Chapter 21 The financial accelerator in a quantitative business cycle framework. *In:* John, B. T. & Michael, W. (eds.) *Handbook of Macroeconomics.* Elsevier.

Borge, L.-E. & J. Rattsø (2008): "Property taxation as incentive for cost control: Empirical evidence for utility services in Norway". *European Economic Review,* 52**,** 1035-1054.

Borgersen, T.-A. & B. K. Kivedal (2012): "Pris-og belåningsgevinster i husholdningenes boligformue". *Praktisk økonomi & finans,* 29**,** 85-91.

Brown, S., M. Dietrich, A. Ortiz-Nuñez & K. Taylor (2011): "Self-employment and attitudes towards risk: Timing and unobserved heterogeneity". *Journal of Economic Psychology,* 32**,** 425-433.

Bryson, A., R. Dorsett & S. Purdon (2002): "The use of propensity score matching in the evaluation of active labour market policies". *Department for Work and Pensions,* working paper no. 4.

Caliendo, M. & S. Kopeinig (2008): "Some practical guidance for the implementation of propensity score matching". *Journal of economic surveys,* 22**,** 31-72.

Cox, N. J. (2006): "WINSOR: Stata module to Winsorize a variable". *Statistical Software Components*.

Dagens Næringsliv. (2015): "Ber Finanstilsynet vurdere tiltak for å dempe boliglånsveksten". Published March 6th 2015. https://www.dn.no/nyheter/naringsliv/2015/03/06/0814/Boligln/ber-finanstilsynet-vurdere-tiltak-for--dempe-boliglnsveksten [Accessed April 21st 2015].

Dehejia, R. H. & S. Wahba (2002): "Propensity score-matching methods for nonexperimental causal studies". *Review of Economics and statistics,* 84**,** 151-161.

DiNardo, J. & J. L. Tobias (2001): "Nonparametric density and regression estimation". *Journal of Economic Perspectives,* 15**,** 11-28.

DiPrete, T. A. & M. Gangl (2004): "Assessing Bias in the Estimation of Causal Effects: Rosenbaum Bounds on Matching Estimators and Instrumental Variables Estimation with Imperfect Instruments". *Sociological Methodology,* 34**,** 271-310.

Disney, R. & J. Gathergood (2009): "Housing wealth, liquidity constraints and self-employment". *Labour Economics,* 16**,** 79-88.

DNB. (2015): "Starting a business: Financing and subsidies". Available: https://www.dnb.no/en/starting-business: [Accessed 13 March 2015].

Ekelund, J., E. Johansson, M.-R. Järvelin & D. Lichtermann (2005): "Self-employment and risk aversion—evidence from psychological test data". *Labour Economics*, 12**,** 649-659.

Ender, P. (2010): "Collin". *STATA Command, UCLA Department of Education*.

Fan, J. (1993): "Local Linear Regression Smoothers and Their Minimax Efficiencies". *The Annals of Statistics,* 21**,** 196-216.

Financial Supervisory Authority of Norway. (2012): "Boliglånsundersøkelsen". Available: http://www.finanstilsynet.no/Global/Bank%20og%20Finans/Banker/Analyser%20og%20statistikk/Boliglaansunders%C3%B8kelsen_hoesten_2012_offentlig_rapport.pdf: [Accessed 12th March 2015].

Fjærli, E., D. Iancu & A. Raknerud (2013): "Facts about entrepreneurship in Norway". Statistics Norway. report No. 52.

Folketrygdloven. (1997): "Lov om folketrygd ". Available: https://lovdata.no/dokument/NL/lov/1997-02-28-19/KAPITTEL_2-1#KAPITTEL_2-1: [Accessed 13 March 2015].

Galiani, S., P. Gertler & E. Schargrodsky (2005): "Water for life: The impact of the privatization of water services on child mortality". *Journal of political economy*, 113**,** 83-120.

Gangl, M. (2004): "RBOUNDS: Stata module to perform Rosenbaum sensitivity analysis for average treatment effects on the treated". *Statistical Software Components*.

Glocker, D. & V. Steiner (2007): "Self-Employment: Way to End Unemployment? Empirical Evidence from German Pseudo-Panel Data". *IZA*, Discussion Paper No. 2561.

Hamilton, B. H. (2000): "Does entrepreneurship pay? An empirical analysis of the returns to self-employment". *Journal of Political economy*, 108**,** 604-631.

Heckman, J. J., H. Ichimura & P. E. Todd (1997): "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme". *The review of economic studies*, 64**,** 605-654.

Heckman, J. J., R. J. LaLonde & J. A. Smith (1999): "The economics and econometrics of active labor market programs". *Handbook of labor economics*, 3**,** 1865-2097.

Heinrich, C., A. Maffioli & G. Vazquez (2010): "A primer for applying propensity-score matching". Inter-American Development Bank. No. IDB-TN-161.

Holland, P. W. (1986): "Statistics and causal inference". *Journal of the American statistical Association*, 81**,** 945-960.

Hurst, E. & A. Lusardi (2004): "Liquidity Constraints, Household Wealth, and Entrepreneurship". *Journal of Political Economy*, 112**,** 319-347.

Hurst, E. & F. Stafford (2004): "Home Is Where the Equity Is: Mortgage Refinancing and Household Consumption". *Journal of Money, Credit and Banking*, 36**,** 985-1014.

IMF. (2009): "Global Financial Stability Report: Navigating the Financial Challenges Ahead". Available: https://www.imf.org/external/pubs/ft/gfsr/2009/02/: [Accessed March 15th 2015].

Innovation Norway. (2015): "Etablerertilskudd". Available: http://innovasjonnorge.no/no/finansiering/etablerertilskudd/#.VQLVQPmjN8E: [Accessed 13 March 2015].

Kaiser, U. & N. Malchow-Møller (2011): "Is self-employment really a bad experience?: The effects of previous self-employment on subsequent wage-employment wages". *Journal of Business Venturing*, 26**,** 572-588.

Koetter, M. & T. Poghosyan (2010): "Real estate prices and bank stability". *Journal of Banking & Finance*, 34**,** 1129-1138.

Lechner, M. (2008): "A Note on the Common Support Problem in Applied Evaluation Studies". *Annales d'Économie et de Statistique***,** 217-235.

Leuven, E. & B. Sianesi (2014): "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing". *Statistical Software Components*.

Lindquist, K.-G. (2012): "Sustainable household debt: Towards an operational view and framework". Norges Bank. Staff Memo, Nr. 33.

Ljunggren, E. (2008): "Kunnskapsbehov om entreprenørskap og kjønn i Norge". *Nordlandsforskning Arbeidsnotat,* 1031/2008.

Mian, A. & A. Sufi (2011): "House Prices, Home Equity-Based Borrowing, and the US Household Leverage Crisis". *American Economic Review*, 101**,** 2132-2156.

NAV. (2015): "Dagpenger under utdanning, opplæring, etablering av egen virksomhet m.v.". Available: https://www.nav.no/rettskildene/Rundskriv/%C2%A7+4-6.+Dagpenger+under+utdanning,+oppl%C3%A6ring,+etablering+av+egen+virksomhet+m.v..354595.cms#Etablering-av-egen-virksomhet: [Accessed March 1st 2015].

NTB. (2015): "Flere banker varsler rentekutt". Available: http://www.dn.no/nyheter/2015/01/04/1527/Renten/flere-banker-varsler-rentekutt: [Accessed March 18th 2015].

OECD. (2004): "Women's entrepreneurship: issues and policies". Available: www.oecd.org/cfe/smes/31919215.pdf: [Accessed March 26th 2015].

Praag, M. v. & A. Raknerud (2014): "The entrepreneurial earnings puzzle". Statistics Norway. Discussion paper No. 789.

Romero, I. & J. A. Martínez-Román (2012): "Self-employment and innovation. Exploring the determinants of innovative behavior in small businesses". *Research Policy*, 41**,** 178-189.

Rosenbaum, P. R. (2002): *Observational studies*, Springer.

Rosenbaum, P. R. & D. B. Rubin (1983): "The central role of the propensity score in observational studies for causal effects". *Biometrika,* 70**,** 41-55.

Rosenbaum, P. R. & D. B. Rubin (1985): "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score". *The American Statistician,* 39**,** 33-38.

Roy, A. D. (1951): "Some thoughts on the distribution of earnings". *Oxford economic papers,* 3**,** 135-146.

Rubin, D. (2001): "Using Propensity Scores to Help Design Observational Studies: Application to the Tobacco Litigation". *Health Services and Outcomes Research Methodology,* 2**,** 169-188.

Rubin, D. B. (1974): "Estimating causal effects of treatments in randomized and nonrandomized studies". *Journal of educational Psychology,* 66**,** 688.

Rønsen, M. (2012): "The family–a barrier or motivation for female entrepreneurship?". Statistics Norway. Discussion paper No. 727.

Smith, H. L. (1997): "Matching with multiple controls to estimate treatment effects in observational studies". *Sociological methodology,* 27**,** 325-353.

Smith, J. (2000): "A critical survey of empirical methods for evaluating active labor market policies". *Research Report, Department of Economics, University of Western Ontario,* No. 2000/6.

Smith, J. A. & P. E. Todd (2005): "Does matching overcome LaLonde's critique of nonexperimental estimators?". *Journal of econometrics,* 125**,** 305-353.

Stambøl, L. S. (2008): "Fordeling og mobilitett av selvstendige næringsdrivende i Norge". *Økonomiske Analyser Statistics Norway,* 2**,** 33-44.

Stambøl, L. S. (2010): "Klarer selvstendig næringsdrivende å opprettholde sin virksomhet over tid?". *Økonomiske Analyser Statistics Norway,* 2**,** 11-20.

Su, X., S. Almaas, L. Bystrøm & F. Carlsen (2014): "Home Equity-Based Refinancing and Household Financial Difficulties: The Case of Norway". *Available at SSRN: http://ssrn.com/abstract=2523025*.

The Research Council of Norway. (2015): "Apply for funding". Available: http://www. forskningsradet.no/en/Apply_for_funding/1138785830985 [Accessed 13 March 2015].

Thurik, A. R., M. A. Carree, A. van Stel & D. B. Audretsch (2008): "Does self-employment reduce unemployment?". *Journal of Business Venturing,* 23**,** 673-686.

Tufte, P. A. (2000): "En intuitiv innføring i logistisk regresjon". *Oslo: Statens institutt for forbruksforskning,* working paper No. 8.

Villund, O. (2005): "Yrkesdata for selvstendig næringsdrivende". Statistics Norway. Dokumentasjonsnotat No. 43.

Vrålstad, S., K. S. Wiggen & L. R. Thorsen (2013): "Levekårsundersøkelsen EU-SILC 2012; Tema: bolig og boforhold, utsatthet og uro for lovbrudd.". Statistics Norway. Dokumentasjonsrapport No. 22.

Wooldridge, J. (2012): *Introductory econometrics: A modern approach*, Cengage Learning.

Zhao, Z. (2003): "Data issues of using matching methods to estimate treatment effects: an illustration with NSW data set". *China Center for Economic Research. ,* Working Paper.

Zhao, Z. (2008): "Sensitivity of propensity score methods to the specifications". *Economics Letters,* 98**,** 309-319.

Appendix A: Summary statistics men and women

Table A.1 Summary statistics of gender difference between the self-employed

	Men N=162			Women N=81		
Variable	Mean	Median	SD	Mean	Median	SD
Cash-Out	0.51	0.37	1.01	0.81	0.52	1.63
Age	47.86	48.00	11.22	45.77	46.00	10.36
Education	0.77	0.50	0.85	0.84	1.00	0.89
Couple	0.88	1.00	0.32	0.89	1.00	0.32
Partner Education	0.23	0.00	0.42	0.23	0.00	0.43
Number of employed persons in household	1.67	2.00	0.69	1.54	2.00	0.67

Note: the number of observations for cash-out is 133 for men and 76 for women.

Appendix B: ATT gender difference tables

Table B.1: Estimated ATT for NN and Caliper matching for men

	Nearest Neighbors			
Algorithm	1	2	5	10
NN Matching without replacement	0.1658 (0.1111) [0 .1392]			
Observations Treated/Untreated	133/133			
NN Matching With replacement	0.2490 (0 .1454)* [0.2202]	0.0894 (0.1231) [0 .1665]	0 .0794 (0.1300) [0.1387]	0 .0620 (0 .1211) [0 .1206]
Observations Treated/Untreated	133/59	133/109	133/239	133/380
Caliper Matching (0.01)	0 .2276 (0 .1488) [0.2401]	0 .0690 (0 .1272) [0 .1667]	0 .0619 (0 .1374) [0 .1533]	0 .0289 (0 .1322) [0.1349]
Observations Treated/Untreated	121/57	121/104	121/231	121/368
Caliper Matching (0.001)	0.2215 (0 .1501) [0.2841]	0.0897 (0.1336) [0 .1931]	0.0538 (0 .1543) [0 .1542]	0 .0190 (0 .1486) [0.1355]
Observations Treated/Untreated	114/52	114/96	114/216	114/333
Caliper Matching (0.0001)	0.2952 (0 .1543)* [0 .2690]	0.1785 (0 .1316) [0.2173]	0 .1214 (0 .1399) [0 .1528]	0 .0802 (0.1323) [0 .1470]
Observations Treated/Untreated	102/49	102/88	102/188	102/284

Standard error in parenthesis, bootstrapped standard error in square brackets.
*** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$

Table B.2 Estimated ATT for Radius, Kernel and Local linear matching for men

	Radius			
Algorithm	0.01	0.005	0.001	0.0005
Radius Matching	-0.0652 (0 .1295) [0 .1230]	-0.0649 (0 .1443) [0 .1256]	-0.0732 (0 .1491) [0 .1352]	-0.0519 (0 .1518) [0 .1558]
Observations Treated/Untreated	121/945	117/911	114/803	112/785
	Bandwidth			
Algorithm	0.1	0.2	0.3	0.4
Kernel matching	-0.0351 (0.1241) [0.1171]	-0.0139 (0.1161) [0.1075]	-0.0153 (0.1092) [0.1072]	-0.0135 (0.1032) [0.1078]
Observations Treated/Untreated	133/987	133/987	133/987	133/987
Local Linear	-0.0257 (0.1454) [0.1244]	-0.0372 (0.1454) [0.0940]	-0.0488 (0.1454) [0.1431]	-0.0753 (0.1454) [0.1240]
Observations Treated/Untreated	133/987	133/987	133/987	133/987

Standard error in parenthesis, bootstrapped standard error in square brackets.
*** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$
For kernel and local linear matching, Epanechnikov kernel function are used.

Table B.3 Estimated ATT for NN and Caliper for women

	Nearest Neighbors			
Algorithm	1	2	5	10
NN Matching without replacement	0 .3992 (0 .2409)* [0 .2352]*			
Observations Treated/Untreated	73/73			
NN Matching With replacement	0.5095 (0.3480) [0.2989]*	0.3598 (0.2639) [0.2672]	0.4516 (0.2270)** [0.2320]*	0.3196 (0.2126) [0.2389]
Observations Treated/Untreated	76/39	73/72	73/162	73/289
Caliper Matching (0.01)	0.5507 (0.3311)* [0.3050]*	0.5443 (0.2673)** [0.2454]**	0.6168 (0.2437)** [0.2299]***	0.5154 (0.2329)** [0.2359]**
Observations Treated/Untreated	64/35	64/68	64/148	64/268
Caliper Matching (0.001)	0.5730 (0.3425)* [0.2955]*	0.5162 (0.2756)* [0.2811]*	0.5521 (0.2500)** [0.2678]**	0.4734 (0.2397)** [0.2219]**
Observations Treated/Untreated	60/33	60/64	60/134	60/209
Caliper Matching (0.0001)	0.5250 (0.3526) [0.3118]*	0.4556 (0.2838) [0.2823]	0.5193 (0.2540)** [0.2791]*	0.4228 (0.2427)* [0.2534]*
Observations Treated/Untreated	57/31	57/58	57/117	57/183

Standard error in parenthesis, bootstrapped standard error in square brackets.
*** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$

Table B.4 Estimated ATT for Radius, Kernel and Local linear matching for women

	Radius			
Algorithm	0.01	0.005	0.001	0.0005
Radius Matching	0.5396 (0.2260)** [0 .2472]**	0.5193 (0.2266)** [0 .2434]**	0.5144 (0.2380)** [0 .2367]**	0 .4720 (0.2360)** [0 .2555]*
Observations Treated/Untreated	64/807	62/732	60/522	58/493

	Bandwidth			
Algorithm	0.1	0.2	0.3	0.4
Kernel matching	0.4299 (0.2061)** [0.2054]**	0.4053 (0.1993)** [0.2051]**	0.3980 (0.1958)** [0.1991]**	0.3845 (0.1948)** [0.2054]*
Observations Treated/Untreated	73/834	73/834	73/834	73/834
Local Linear	0.4546 (0.3480) [0.1988]**	0.4294 (0.3480) [0.2501]*	0.4057 (0.3480) [0.2097]*	0.4017 (0.3480) [0.2092]*
Observations Treated/Untreated	73/834	73/834	73/834	73/834

Standard error in parenthesis, bootstrapped standard error in square brackets.
*** = $p < 0.01$, ** = $p < 0.05$, * = $p < 0.1$
For kernel and local linear matching, Epanechnikov kernel function are used.

Appendix C: Winsorize

There are several options in dealing with outliers. The first alternative is to ignore the problem, but this can heavily affect the distribution and give biased estimators. Another variant is to trim the data, where you clip of some percentile of the top and/or bottom of the distribution. This may be effective but does in a way censor data, by removing some of the distribution. An alternative to trimming the data is to winsorize it, where you do not remove or censor any data but instead do a transformation of the extreme values over some percentile.

When winsorizing, the data is ordered (not including missing data) in such a way that observations (x):

$$x_1 <= \ldots <= x_n$$

Then new variables (y) are created which are identical to x except that the lower and/or higher h variables are replaced by the next value counting inward from the extremes:

$$y_1, \ldots, y_h = y_{h+1}$$

$$y_n, \ldots, y_{n-h+1} = y_{n-h}$$

The specification of h can either be a percentile or any specified number of variables counting from the extremes.

In figure C.1 we see the result of winsorizing the cash-out variable, accomplished by using the Stata module WINSOR (Cox, 2006). Before winsorizing there are long thin tails in the distribution, which is transformed after winsorizing.

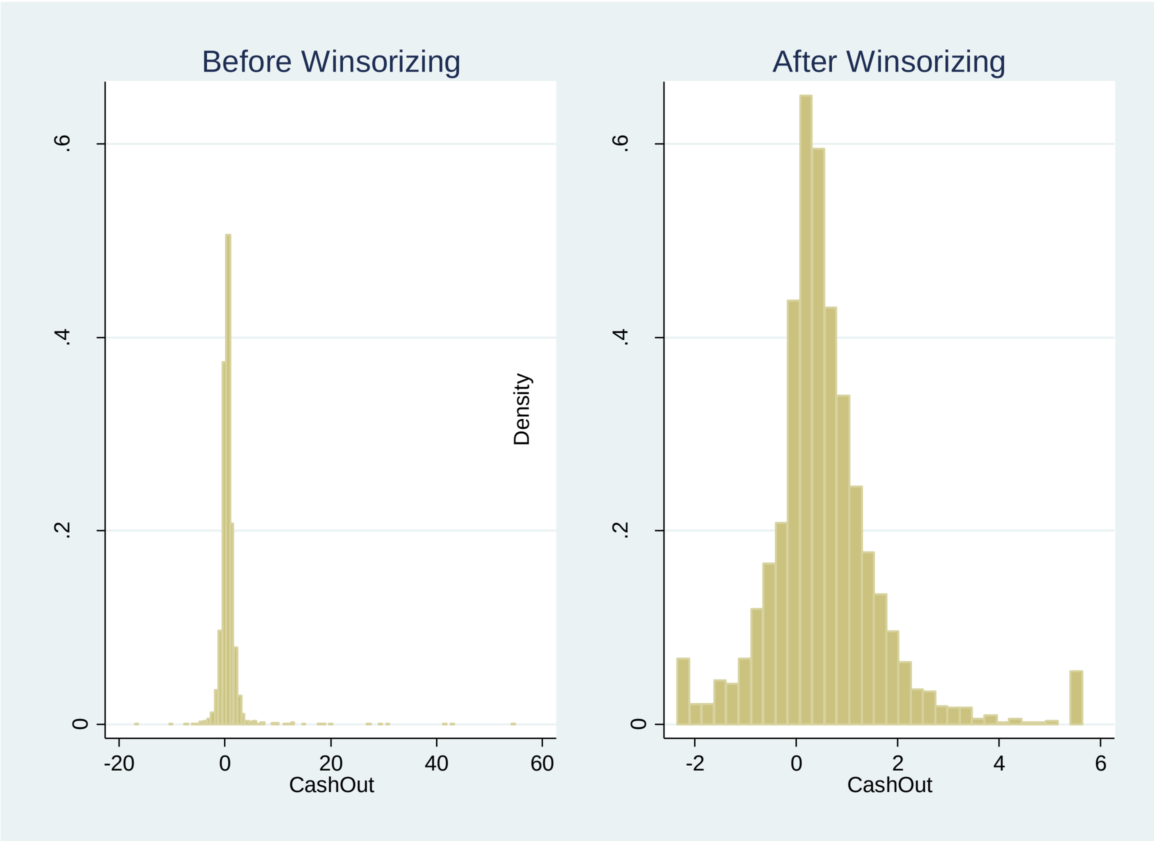

Figure C.1 Distribution of cash-out before and after winsorizing at 1st and 99th percentile
Note: different scale on x-axis

Appendix D: Matching quality trimmed sample

Table D.1 Bias reduction in the control variables within trimmed sample

Variable	Unmatched Matched	Mean		%Bias	%Reduced Bias	t-test	
		Treated	Untreated			t	P > |t|
Age	U	5.7703	5.1697	28.4		3.88	0.000
	M	5.8095	5.6508	7.5	73.6	0.61	0.543
AgeSquared	U	37.722	31.233	26.6		3.77	0.000
	M	38.127	36.032	8.6	67.7	0.69	0.491
Gender	U	0.6363	0.5420	19.2		2.60	0.009
	M	0.5714	0.5793	-1.6	91.6	-0.13	0.899
Education	U	0.7799	1.3185	-67.1		-9.66	0.000
	M	1.1270	1.1032	3.0	95.6	0.23	0.815
Couple	U	0.8899	0.8286	17.7		2.26	0.024
	M	0.8412	0.8571	-4.6	74.1	-0.35	0.726
PartnerEducation	U	0.2248	0.0335	59.4		12.08	0.000
	M	0	0	0.0	100.0	.	.
NumEmployed	U	1.6746	1.7814	-16.8		-2.37	0.018
	M	1.7460	1.7937	-7.5	55.4	-0.67	0.506

The table presents mean values of treated and untreated before and after matching for each control variable. %Bias is the standardized bias, and %Reduced Bias is the reduction in standardized bias after matching. t values are reported from test for difference in means of treated and untreated.

Table D.2 Difference between unmatched and matched within trimmed sample

Sample	Pseudo R Squared	Likelihood Ratio chi2	P > chi2	Mean Bias	Median Bias	Rubin's B	Rubin's R
Unmatched	0.142	191.10	0.000	33.6	26.6	96.5	2.29
Matched	0.003	0.93	0.988	4.7	4.6	12.1*	1.69*

The table presents the difference of Pseudo R squared, mean bias, median bias and Rubin's R and B for unmatched and matched sample.
* if B < 25%, R in [0.5, 2]

Appendix E: Test of multicollinarity

We test for multicollinarity by utilizing the Stata module COLLIN (Ender, 2010), which reports the variance inflation factor (VIF) for every control variable. VIF is determined by the correlation of x_i and the other explanatory variables.

$$VIF_i = \frac{1}{(1 - R_i^2)}$$

Where R_i^2 is the *R-squared* from regressing x_i on all other explanatory variables including a constant term (Woolridge, 2012). The VIF's for the variables included in our logit-model, is presented in table E.1. Inclusion of the squared age variable, gives multicollinarity because of how squared terms are likely to be correlated with its root. Such cases of multicollinarity are of no harm since they have minimal effect on the p-values. By excluding *AgeSquared* we get VIF values below 10 for all control variables. We conclude that there is no sign of strong multicollinarity in our logit model.

Table E.1 Variance Inflation Factor

	Variance Inflation Factor	
Variable	Inclusion of Age squared	Exclusion of Age squared
AgeSquared	17.63	-
Age	17.44	1.01
Gender	1.01	1.01
Education	1.03	1.02
Couple	1.26	1.26
PartnerEducation	1.04	1.03
NumEmployed	1.29	1.25

Appendix F: Matching estimators

This appendix serves as an extension of chapter 5.5, where we give a brief mathematical representation of matching estimators as given by Smith and Todd (2005). The ATT matching estimator have the following form:

$$\hat{\tau}_{ATT}^{PSM} = \frac{1}{N_1} \sum_{i \in I_1 \cap S_P} \left(Y_{1i} - \sum_{j \in I_0} W(i,j) Y_{0j} \right) \tag{F.1}$$

S_P defines the region of common support, I_1 and I_0 are the set of treated and untreated respectfully and N_1 is the number of observations in set $I_1 \cap S_P$. $W(i,j)$ is a weighting function dependent on the distance between the treated propensity score (P_i), and the untreated propensity score (P_j). The variance of the ATT estimator, as estimated by PSMATCH2, is given by:

$$Var(\hat{\tau}_{ATT}^{PSM}) = \frac{1}{N_1} Var(Y_1 | D = 1) + \frac{1}{N_1^2} \sum_{j \in I_0} W(i,j)^2 \, Var(Y_0 | D = 0) \tag{F.2}$$

$C(P_i)$ defines a propensity score neighborhood around each treated with neighbors form the untreated propensity score sample. Individuals j matched to treated individual i are those individuals in set $A_i = \{ j \in I_0 | P_j \in C(P_i) \}$. NN matching defines $C(P_i) = \min_i \| P_i - P_j \|$, $j \in I_0$, and caliper matching defines $C(P_i) = \{ P_j \mid \| P_i - P_j \| < \epsilon \}$, where ϵ is the restricting caliper. Oversampling in these algorithms in our thesis is done with uniform weighting.

Kernel and local linear defines $C(P_i) = \left\{ \left| \frac{P_i - P_j}{a_n} \right| \leq 1 \right\}, j \in I_0$, where a_n is a bandwidth parameter. The weight in kernel matching is given by:

$$W(i,j) = \frac{K\left(\frac{P_j - P_i}{a_n}\right)}{\sum_{k \in I_0} K\left(\frac{P_k - P_i}{a_n}\right)} \tag{F.3}$$

Where K is a kernel function (Epanechnikov kernel in our thesis) with the form $K(x) = \frac{3}{4}(1 - x^2)\mathbf{1}_{|x| \leq 1}$, where $\mathbf{1}_{|x| \leq 1}$ is an indicator function restricting the function values to only non-negative outcomes. The counterfactual outcome for kernel matching, can be viewed as the solution to the $\hat{\alpha}$ estimator from the weighted regression on an intercept: $\min_{\alpha} \sum_{j \in I_0} \left(Y_{0j} - \alpha\right)^2 K\left(\frac{P_j - P_i}{a_n}\right)$.

In local linear the weight is given by:

$$W(i,j) = \frac{K_{ij} \sum_{k \in I_0} K_{ik}(P_k - P_i)^2 - [K_{ij}\left(P_j - P_i\right)][\sum_{k \in I_0} K_{ij}(P_k - P_i)]}{\sum_{j \in I_0} K_{ij} \sum_{k \in I_0} K_{ik}(P_k - P_i)^2 - \left(\sum_{k \in I_0} K_{ik}(P_k - P_i)\right)^2} \qquad \text{(F.4)}$$

Where $K_{ij} = K((P_j - P_i)/a_n))$. Fan (1993) shows that the local linear estimator for the counterfactual outcome can be viewed as the solution to the $\hat{\alpha}$ estimator from the weighted regression with an intercept and a slope: $\min_{\alpha,\beta} \sum_{j \in I_0} \left(Y_{0j} - \alpha - \beta\left(P_j - P_i\right)\right)^2 K\left(\frac{P_j - P_i}{a_n}\right)$.