University of Essex

Doctoral Thesis

Essays in empirical microeconomics and finance

Author:
Stefano Alderighi

Supervisors:
Prof. Eric Smith
Prof. Sule Alan

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Department of Economics

May 2016

"Try to understand men. If you understand each other you will be kind to each other. Knowing a man well never leads to hate and almost always leads to love."

John Steinbeck

Essays in empirical microeconomics and finance

by Stefano ALDERIGHI

Summary

The present thesis is divided in three chapters. The first focuses on Education Economics. The second and the third on Household Finance. The following paragraphs describe the contents of each chapter more in detail.

The first essay compares and contrasts aggregate and individual level analyses to investigate the relationship between economic fluctuations and tertiary education enrollments in Italy. It shows that aggregate enrollments follow a procyclical pattern. Consistently, it finds that Italian young individuals living with their parents implement a procyclical enrollment decision. The paper tries to reconcile the empirical evidence with theoretical predictions, and investigates a number of different channels. It proposes a rather novel, nevertheless theory consistent interpretation for the evidence found. The paper argues that Italian individuals living with their parents implement a procyclical behaviour because they can't access credit to finance their education. It supports this statement with consistent empirical evidence.

The second essay studies whether labour income volatility crowds out investment in risky assets in Italian households. Justified by the literature on limited participation, the paper makes use of reduced form estimations to show that Italian households hedge their labour income risk on the risky assets market. It contributes by proposing a novel measure of labour income volatility, ground on the literature on labour income dynamics. On a methodological perspective, the paper adds to the literature by estimating reduced form models using a recent estimation technique, never implemented before in this branch of research. It shows that the new methodology overcomes some of the limitations of the techniques previously applied in the literature.

The third essay, co-authored with Professors Sule Alan and Eric Smith, focuses on the subprime credit market in the United Kingdom. Making use of a unique database centered on a randomized trial experiment, the paper identifies the causal effect of an increase in the cost of credit on individual credit demand and default probability.

Acknowledgements

It goes without saying that I am immensely grateful to my supervisors: Professor Eric Smith, who dedicated time and effort to teach me the profession, making me a (hopefully) good researcher and a (hopefully) more focused person; and Professor Sule Alan, who taught me exactly the same things as he did, but following a different (and complementary) perspective. A massive thanks also goes to my chair, Doctor David Reinstein, who has constantly given me useful comments and has believed in my research from the beginning.

Other members of staff have been very patient and helpful along these years. Professor Joao Santos Silva, who always had a minute to talk and advise; Professor Marco Francesconi, whom I thank for his constant support; Professor Gianluigi Vernasca, who made me a good teacher and a better economist.

However, I must admit that without my friends I could never have made it. They are many, fortunately. I will not name them here: they know who they are. I thank those who were already here, and eventually left. I thank those who started with me, supported me every day, and are finishing with me. I thank those who arrived after me, and will remain here once I will be moving away.

One person is worth (and due) explicitly mentioning. She is my partner, and my best friend. My thesis, and whatever good I am doing, is dedicated to her. Thank you Silviya for sharing this path with me. I would be happy if I am making you one tenth of the good person you made me become.

Contents

Abstract ii

Acknowledgements iii

Contents iv

1 Introduction. 1
- 1.1 Justification of the present work . . . 1
- 1.2 Overview . . . 2

2 The Cyclicality of University Enrollments in Italy. 5
- 2.1 Introduction . . . 5
- 2.2 Literature review . . . 8
- 2.3 Stylized facts in the Italian economy . . . 10
- 2.4 Database description, sample selection and descriptive statistics . . . 11
  - 2.4.1 Database description . . . 11
  - 2.4.2 Sample selection . . . 12
  - 2.4.3 Descriptive statistics . . . 14
- 2.5 Empirical results . . . 16
  - 2.5.1 The estimation techniques . . . 16
  - 2.5.2 The model specification . . . 17
  - 2.5.3 Main empirical findings: first year enrollment decision . . . 18
  - 2.5.4 First year enrollment decision and labour market outcomes. . . . 22
  - 2.5.5 The enrollment decision at any year. . . . 25
    - 2.5.5.1 Pooled logistic regressions . . . 26
    - 2.5.5.2 Enrollment decision (at any year) and labour market outcomes . 28
    - 2.5.5.3 Fixed-effects . . . 29
  - 2.5.6 Reconciling the findings with the theory . . . 31
    - 2.5.6.1 Household borrowing constraints . . . 32
    - 2.5.6.2 Future labour market conditions . . . 35
    - 2.5.6.3 Reverse causality . . . 36
    - 2.5.6.4 Supply side considerations . . . 38
    - 2.5.6.5 Individuals living on their own . . . 40
- 2.6 Conclusions . . . 41

3 Labour Income Risk Hedging in Italian Households. 49
3.1 Introduction . . . 49
3.2 Literature review . . . 51
3.3 Database description and sample selection . . . 53
3.4 Labour income risk score . . . 54
3.4.1 Theoretical background . . . 54
3.4.2 The creation of a measure of idiosyncratic risk . . . 56
3.5 Estimations . . . 58
3.5.1 Theoretical background . . . 58
3.5.2 The estimation technique . . . 59
3.5.3 The model specification . . . 61
3.5.4 Descriptive statistics . . . 62
3.5.5 Estimation results . . . 64
3.5.5.1 Main results . . . 64
3.5.5.2 Robustness check: alternative techniques . . . 67
3.5.5.3 Robustness check: alternative dependent variables . . . 69
3.5.5.4 Robustness check: risk distribution within the household . . . 71
3.5.5.5 Small sample size . . . 72
3.5.6 Endogeneity . . . 73
3.5.6.1 Attitude towards risk . . . 74
3.5.6.2 Bargaining power . . . 75
3.5.6.3 Covariance income shocks - risky assets returns . . . 76
3.6 Conclusions . . . 77

4 Interest Rates and Default in Subprime Credit Markets: Evidence from a Randomized Price Experiment (With Sule Alan and Eric Smith). 91
4.1 Introduction . . . 91
4.2 Theoretical motivations . . . 95
4.2.1 Constrained individuals . . . 98
4.2.2 Outcome variables . . . 99
4.3 Data and Experimental Design . . . 101
4.3.1 The lender . . . 101
4.3.2 The experimental design . . . 101
4.3.3 Data and descriptive statistics . . . 103
4.3.4 Internal validity tests . . . 104
4.3.5 External validity . . . 105
4.4 Results . . . 106
4.4.1 Empirical Model . . . 106
4.4.2 Main empirical results . . . 107
4.4.3 Interest rate sensitivity of returns . . . 110
4.4.4 Robustness check: transitions . . . 112
4.4.5 Robustness check: duration models . . . 114
4.4.6 Access to other sources of credit . . . 115
4.5 Conclusions . . . 117

Bibliography **119**

Someone wrote: "To my folks". I quote.

Chapter 1

Introduction.

1.1 Justification of the present work

"Like any other science, economics
is not content with merely descriptive knowledge.
It tries to discern general patterns of uniformity
in the administration of scarce resources"

[Lange, 1945].

The present doctoral thesis contains three pieces of Applied Economics research. The first of them is focused on educational choices. The remaining two on household and individual financial decisions. The above quotation [Lange, 1945] represents an ideal starting point to justify the choice of applied economics as a *trait d'union* for the present doctoral thesis. The sentence summarizes some features of economics as a discipline, and gives information about its method of investigation. Firstly, it defines economics as a science ("like any other science"). More in detail, Lange defines economics as an *empirical science*, on the ground that its assumptions are "approximate generalization of empirical observations" [Lange, 1945, page 21]. Secondly, it suggests that economics is not and cannot be a display of empirical evidence collected for mere descriptive purposes. It is a deductive discipline, with theoretical predictions that need empirical validation [Lange, 1945, page 21].

In economics, theoretical predictions often postulate causal relations among or between variables. Under this assumption, it is clear that a successful empirical validation should coincide, or be the closest possible to, causal inference. As Hoover [2005] mentions, a natural way to empirically validate a causal relationship is to look for counterfactual evidence to what the theory predicts. However, is it easy to find counterfuctual evidence in economics, and in general in social sciences? Hayek [1967] describes social phenomena as 'complex', as opposed to phenomena in natural sciences. By 'complex', Hayek means that social phenomena are driven by a large and often not countable (or predictable) amount of variables. Given this composite influences, as a matter of fact isolating counterfactuals and identifying causal effects in economics is not a trivial task [Addison et al., 1984].

The difficulty in identifying counterfactuals is not the only impediment to causal inference. A related problem concerns how assumptions and theorems are derived in economics. Firstly, economics assumptions are often approximation of empirical regularities. As a consequence, the two are characterized by a physiological mismatch [Lange, 1945]. Secondly, theorems based on economics assumptions are generally the result of manipulations that go beyond or neglect empirical evidence [Lange, 1945]. As a consequence, they might significantly deviate from what happens in the real world.

Do the three chapter address some, or even all of the above points? The following section aims at describing the three chapter more in detail, and at defending them as 'well behaved' pieces of Applied Economics research.

1.2 Overview

The thesis is divided in three chapters.

The first chapter (chapter 2, in the present thesis) investigates the cyclical behaviour of Italian tertiary education enrollments. Economic theory predicts that, in absence of borrowing constraints, human capital accumulation should in principal be countercyclical because of opportunity cost considerations [Dellas and Sakellaris, 2003]. In a recession, for example, the opportunity cost of acquiring education is low: both the wage rate and the probability of finding a job (typically procyclical variables) are negatively affected by the business cycle. However, if a recession today is expected to persist and negatively influence future economic conditions, individuals might decide to acquire *less* education, and implement a procyclical enrollment decision [Micklewright et al., 1990]. Italian aggregate and individual data show that individuals aged 18-25 and living with their parents implement a procyclical enrollment decision. Although at odds with opportunity cost considerations, this finding is consistent with the presence of borrowing constraints at a household level, and with the fact that individuals might expect bad (or good) economic conditions to persist in the future. These two channels, however, are not supported by the data. Further evidence shows that borrowing constraints might operate at an individual level. Young individuals living on their own implement indeed a countercyclical enrollment decision, consistently with opportunity cost considerations. Among these individuals, the most financially exposed (whom I consider borrowing constrained) implement instead a procyclical enrollment choice.

Reconnecting this summary with the previous section, it is clear how the decision of enrolling to university is a 'complex' phenomenon [Hayek, 1967]. Regarding how individuals should behave when facing economic fluctuations, economic predictions are rather straightforward. Yet, the reason why the evidence found apparently deviates from the theory is not immediately clear. The paper can be considered as a good piece of applied research, as it provides robust evidence of a finding (procyclical enrollments), addresses several channels the literature identifies to explain this phenomenon, and provides a novel but theoretically consistent explanation of the evidence.

The second chapter (chapter 3, in the present thesis) assesses whether uninsurable labour income volatility negatively influences household investment in risky assets showing positive expected returns. The paper shows that this is the case: consistently with theoretical predictions [Gomes and Michaelides, 2005], a higher labour income risk significantly reduces household investment in risky assets. The result is robust the the introduction of an increasing number of regressors, to inference based on a bootstrapped variance covariance matrix, to the implementation of different techniques and different dependent variables. As labour income risk is an endogenous regressor [Betermier et al., 2012], the paper also addresses the endogeneity problem. It provides suggestive evidence that labour income risk has a direct influence on investment in risky assets, and that this is due to hedging reasons. The paper contains two more contributions. The first lies in the creation of a novel measure of labour income risk. The second, methodological, refers to the implementation of a novel, and more appropriate, estimation technique.

The problem herein addressed is mainly that of the mismatch between baseline theoretical predictions and empirical regularities [Lange, 1945]. Under the complete markets assumption, economic theory predicts indeed that everyone should invest a positive (and relevant) fraction of their wealth in risky assets. As it happens not to be the case (the 'stockholding puzzle': Haliassos and Bertaut [1995]), economics provides several explanations for this mismatch. Among them, uninsurable labour income volatility. The paper performs well in providing consistent and supporting evidence towards this explanation. Albeit not able to isolate a causal effect, it addresses this problem, and performs a number of robustness checks that suggest the presence a direct link between labour income risk and investment in risky assets. The creation of a novel measure of labour income risk also moves towards this direction, as the paper proposes an indicator that is consistent with economic theory but also computationally efficient.

The third chapter (chapter 4, in the present thesis) exploits experimental data to isolate a causal effect between the cost of credit on credit demand and individual default probability. Variations in the cost of credit are typically endogenous to borrowers' characteristics: riskier individuals should in principle be charged higher interest rates Stiglitz and Weiss [1981]. As a matter of fact, only an experimental design can allow the observation of a counterfactual in this situation. The paper implements a propriety database with such design to draw causal inference on this issue. It checks whether the experiment has statistical validity, and provides robustness checks in support of the main evidence.

Reconnecting this summary with the previous section, this paper does a good job as an Applied Economics paper as it isolates a valid counterfactual [Hoover, 2005], and provides empirical evidence that is consistent with economic theory.

Chapter 2

The Cyclicality of University Enrollments in Italy.

Summary

The present chapter compares and contrasts aggregate and individual analyses to study whether Italian tertiary education enrollments are influenced by the business cycle. At odds with most empirical evidence, it firstly shows that aggregate enrollments follow a procyclical pattern. Individuals analyses, performed on a sample of young individuals living with their parents, confirm this finding. The paper explores several channels to reconcile these results with theoretical predictions. It argues that the pattern can be attributed to individual borrowing constraints. Most young Italian individuals living with their parents do not have access to unsecured debt. As a consequence, the funding of their tertiary education heavily depends on household resources, and on their fluctuations.

2.1 Introduction

In the last ten years, the Italian press has been engulfed with scaremonger articles on how University enrollments have been steadily declining, and on how this fact is attributable to the 2007 crisis and to the scar it left on the Italian economy.[1] More in detail, Italian newspapers explain that households hit by the Great Recession (and never fully recovered) are not able to afford university tuition fees and other educational expenses. As a consequence, they prefer their children not to enroll in tertiary education.

[1] Links to document this tendency of the Italian press are available upon request. All articles are in Italian.

Italian newspapers are implicitly stating that enrollments are positively correlated with economic fluctuations. Press articles are not rigorous academic studies. However, if the public opinion and major newspapers (from all political views) unanimously agree on this issue, it is likely that their statements have some foundations. Economic theory is apparently at odds with what the Italian press reports. During economic downturns the opportunity cost of acquiring education is reduced. Both the wage rate and the probability of finding a job are indeed low. As a consequence, during a recession enrollments should increase (consistently with a countercyclical schooling decision: Dellas and Sakellaris [2003]). This result holds in absence of borrowing constraints. In Italy tuition fees are relatively low, and students are characterized by scarce mobility (Brunello and Cappellari [2008], Staffolani and Pigini [2012]). As a consequence, it is reasonable to assume that the average Italian household is not credit constrained for tertiary education expenses. Opportunity cost considerations would intuitively lead to a countercyclical pattern in Italian enrollments.

The opportunity cost of studying is not the only driving force of the relationship between schooling choices and the business cycle [Micklewright et al., 1990]. Economic fluctuations can indeed also influence the expected returns to schooling, and in general expectations on future economic conditions. If today's downturn is likely to negatively influence future economic prospects, during a recession individuals can be discouraged to acquire further education (Tumino and Taylor [2015], Petrongolo and San Segundo [2002]). The relationship between schooling decisions and economic fluctuations cannot be determined *a priori*, and needs empirical investigation to be fully understood. The paper aims at describing whether the pattern depicted by Italian newspapers is present in the data, and at providing an explanation for the evidence found.

The paper is an empirical contribution. It firstly analyses Italian macroeconomic data, and shows that aggregate enrollments follow a procyclical pattern. The result is robust to different enrollment measures. The paper then turns to the analysis of individual data. It makes use of an Italian longitudinal database (SHIW, Bank of Italy) to estimate reduced form specifications. It firstly shows that the enrollment decision shows a procyclical pattern *even* at an individual level. It then explores different channels to reconcile the empirical findings with the theory. It does not find strong evidence of household borrowing constraints (similarly to Christian [2007]). Moreover, it cannot explain the procyclical pattern through the relationship between current and future economic conditions (finding different results from Tumino and Taylor [2015] and Petrongolo and San Segundo [2002]). It argues instead that borrowing constraints operate at an individual level. Indeed, young Italian individuals living with their families do not have easy access to credit to fund their education, and strongly depend on household resources to continue their studies. Consistently, young individuals living on their own show a strongly countercyclical enrollment decision. This result reconciles the findings with economic theory and most of the empirical evidence on the topic.

The paper contributes to the literature in several ways. It adds empirical evidence to a small, empirical literature on the cyclicality of educational choices (as so defined by Méndez and Sepúlveda [2012]: Betts and McFarland [1995], Dellas and Sakellaris [2003], Christian [2007], Sakellaris and Spilimbergo [2000], Alessandrini [2014]). It relates this literature to a parallel one, focused more in general on the relationship between economic conditions and schooling decisions (Pissarides [1981], Pissarides [1982], Micklewright et al. [1990], Petrongolo and San Segundo [2002], Tumino and Taylor [2015]).

The paper is one of the few on the topic that explicitly focuses on a continental European country. Most of the cited contributions concentrate indeed on North America or on the UK (Tumino and Taylor [2015], Pissarides [1981], Pissarides [1982] and Micklewright et al. [1990] on the UK, Alessandrini [2014] on Canada, the rest of the studies on US). Petrongolo and San Segundo [2002] represents an exception to this trend. Méndez and Sepúlveda [2012] express concern on this issue. They attribute the lack of agreement of previous research to institutional differences across countries. This paper supports their concern and broadens the view on the topic, as the pattern emerged in the data is eventually explained by invoking cultural, institutional and regulation differences between Italy and more free-market oriented economies.

On a methodological ground, the paper embraces a wide range of techniques previously used in similar research questions. It exploits substitutability between labour market activities and human capital acquisition to estimate a joint model of employment and enrollment decision. Similarly to Dellas and Sakellaris [2003] and Petrongolo and San Segundo [2002], and following the theoretical implications of Perli and Sakellaris [1998], it investigates the mentioned substitutability by estimating a multinomial logit regression. Similarly to Méndez and Sepúlveda [2012], it makes use of panel data to control for unobserved individual heterogeneity. To explain the findings emerged from the data, the paper explores a number of theoretically driven and empirically supported explanations (as in Christian [2007], Tumino and Taylor [2015]). It adds to the literature by providing a novel interpretation to the empirical results.

Studying the cyclicality of university enrollments is not just an exercise to see whether empirical regularities match with economic theory. The cyclical behaviour of skill acquisition has fundamental macroeconomic implications. Perli and Sakellaris [1998], motivated by Dellas and Sakellaris [2003], introduce skill acquisition in an RBC framework, and show that human capital formation influences the persistence of economic growth. DeJong and Ingram [2001] follow this line of research. They show that skill acquisition is countercyclical, and that in turn it influences business cycle fluctuations. Reconstructing what macroeconomic consequences arise from reduced form results can be hazardous. However, given the theoretical predictions, a countercyclical enrollment decision would have desirable long run consequences.[2] Under this assumption, the explanation proposed to the empirical findings has a straightforward policy implication: release young individuals from their 'family credit constraints'. This outcome can be achieved by enhancing students' opportunities to participate in the labour market, or by simply allowing them to borrow money to finance their education.

The rest of the paper is organized as follows: section 2.2 contains a literature review. Section 2.3 outlines stylized facts in the Italian economy. Section 2.4 describes the database and justifies the sample selection. Section 2.5 contains the estimation results. Section 2.6 concludes.

2.2 Literature review

The literature is giving increasing attention to the cyclicality of schooling decisions. Early contributions analyze how educational choices relate to "economic conditions", without explicitly taking cyclicality into consideration (Pissarides [1981], Pissarides [1982], Mattila [1982], Micklewright et al. [1990]). Pissarides [1982] is of particular interest for this paper, as it studies early school leaving and university enrollment choices. The paper suggests that schooling choices should have a countercyclical pattern, and eventually finds that an increase in adult unemployment positively influences the staying-on rate of 16 years old students. This has in turn long run positive consequences on university enrollment decisions. Betts and McFarland [1995] and Dellas and Sakellaris [2003] are the first empirical contributions explicitly focusing on the cyclicality of schooling decisions. Betts and McFarland [1995] study the cyclical behaviour of U.S. community colleges enrollments. Dellas and Sakellaris [2003] study instead the cyclicality of the demand for higher education in the U.S.. They both find that schooling decisions are countercyclical. Betts and McFarland [1995] find that a one-percent increase in the unemployment rate increases full-time attendance in public-two year colleges by 4.5%. Dellas and Sakellaris [2003] find that a one percent increase in the unemployment rate is associated with an increase in enrollments by 2%.

After these contributions, the attention towards the topic has been steadily growing. Christian [2007] analyses the role of borrowing constraints in explaining U.S. college enrollments. He finds mixed evidence on the topic. He shows that enrollments of individuals in (expected) low income

[2]As DeJong and Ingram [2001] emphasize, investment in physical capital is generally procyclical. The accumulation of physical and human capital balance each other along fluctuations, ensuring that agents are able to smooth their consumption [DeJong and Ingram, 2001]. When a positive technology shock hits the economy, both low skilled and high skilled labour are demanded more [Perli and Sakellaris, 1998]. The low elasticity of substitution between high and low skilled labour in the skill acquisition sectors ensures that the release of high skilled labour towards the production sector happens slowly. This mechanism ensures a stable and durable output growth pattern.

families are significantly more procyclical than those of individuals in (expected) higher income ones. This result is not confirmed when he compares house owning and non-owning households.

As Dellas and Sakellaris [2003] mention, formal education is not the only form of human capital acquisition worth investigating. Méndez and Sepúlveda [2012], for example, study the cyclical behaviour of both secondary schooling and job-specific training. They find that aggregate enrollments present relevant composition effects, and conclude that the cyclicality of school acquisition differs across categories. Employed individuals acquire training procylically, whilst unemployed individuals implement an opposite choice. Firm financed training programs are procyclical, whilst self financed programs are countercyclical. Alessandrini [2014] analyses post-compulsory schooling decisions in Canada. She finds that college enrollments are procyclical and university enrollments are countercyclical. She suggests that downturns stimulate the accumulation of general human capital, and booming periods push the acquisition of more technical skills.

Similarly to Pissarides [1981] and Pissarides [1982], some recent contributions focus on the relationship between schooling decisions and labor market conditions, without explicitly mentioning cyclical fluctuations. Petrongolo and San Segundo [2002] study whether local Spanish labour market indicators *and* family background influence the staying-on decision of 16 years old individuals. They find that young unemployment rate is positively correlated with the staying-on decision, consistently with opportunity cost considerations. They also show that adult unemployment rate has a negative influence on the staying-on choice, consistently with an expected-returns explanation [Micklewright et al., 1990]. They emphasize, however, that although labour market conditions matter, the family background of the students is the main driving force of their results.

Tumino and Taylor [2015], motivated by the Great Recession, analyze how labour market conditions influence school leaving decisions for British 16 years old individuals. Albeit they do not explicitly focus on the cyclicality of schooling choices, they investigate how young and adult unemployment rates relate to the staying-on decision after compulsory schooling. They proxy the opportunity cost of studying with young unemployment rate, and the expected future economic conditions with adult unemployment rate (following Petrongolo and San Segundo [2002]). They find that individuals from low socio-economic background continue their studies when young unemployment rate is higher, implementing a countercyclical enrollment decision.

2.3 Stylized facts in the Italian economy

The present section analyses how first year enrollments move along the business cycle, using Italian aggregate data.

The following graph plots first year enrollments (in units, ISTAT) against two macroeconomic indicators. The first is employment level in the 15-24 class age (thousands of unit, OECD). The second is the unemployment level in the 15-24 class age (thousands of unit, OECD).[3] All variables are detrended using an Hodrick Prescott filter ($\lambda = 100$).[4] The time span chosen is 1990-2010.[5] Both graphs suggest that first year enrollments in tertiary education have a procyclical behaviour. The result is at odds with opportunity cost considerations.

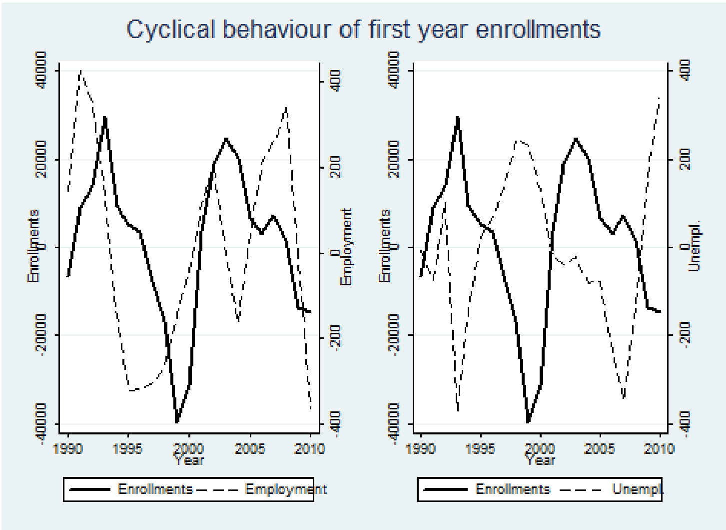

FIGURE 2.1: hp filtered variables

[3]Unless explicitly stated, the following analyses will use the unemployment indicator as a measure of business cycle fluctuations.

[4]The literature hasn't found an agreement on the magnitude of the λ parameter [Maravall and del Río, 2001]. The paper follows Giorno et al. [1995] in setting the parameter equal to 100.

[5]This period is chosen to match aggregate and individual data (see section 2.4). The same analysis, performed on the 1977-2010 time span allows to draw similar conclusions. The results are displayed in appendix 1 (Table 2.17)

Although graphical evidence is informative, it does not allow to quantify this relationship. The following table displays OLS regressions of first year enrollments on the two macroeconomic measures introduced above, and a number of other indicators.[6] All variables (including the dependent one) are expressed in logarithms, and are detrended using an Hodrick-Prescott filter ($\lambda = 100$). As a consequence OLS coefficients can be interpreted as elasticities.

The regression results confirm the evidence displayed in the graph. First year enrollments in tertiary education have a procyclical behaviour, at odds with opportunity cost considerations. When unemployment (in the 15-24 class age) is above the cycle by 1%, enrollments decrease by 1.84%. The remaining coefficients have a similar interpretation. Estimation making use of rates rather than de-trended macroeconomic indicators leads to the same conclusions. The results are displayed in appendix 1, table 2.18.

The evidence displayed is rather unique among the empirical findings. Dellas and Sakellaris [2003] propose a similar analysis for the U.S.. The authors plot detrended unemployment rates against detrended college enrollment rates, finding a countercyclical pattern. They also perform a regression analysis, and they show that an increase in the unemployment rate by one percentage point is associated with an increase in the enrollment rate by 0.57 percentage points. Alessandrini [2014], focuses on post-secondary education in Canada. She shows that enrollment rates, plotted against GDP, have a countercyclical pattern, and she supports this finding with a correlation analysis. Méndez and Sepúlveda [2012] plot schooling against unemployment rate, and emphasize a positive and significant correlation between the two variables. Betts and McFarland [1995] introduce a similar graph, in which U.S. community college enrollments positively relate with the unemployment rate, suggesting a countercyclical behaviour.

2.4 Database description, sample selection and descriptive statistics

2.4.1 Database description

The data are taken from the Survey of Household Income and Wealth of the Bank of Italy (SHIW from now on). The SHIW database is a random sample of nearly 8,000 households per year. Although it is available from 1977 it shows an actual panel structure only from 1989 onwards. Due to attrition between 1989 and 1991, the analyses proposed pertain to the 1991 - 2010 time span, and make use of ten waves in total.

[6] Unemployment, employment, and self-employment are national variables, in levels, measured in thousands of units (ISTAT). GDP (OECD) and GDP (ISTAT) are alternative measures of GDP, in levels, PPP, at 2005 prices (in millions of dollars).

The database is suitable to study the proposed question from a micro perspective. Although it does not contain a variable that explicitly states whether an individual is enrolled in university or not, it allows the construction of such an indicator. Firstly, it details about the maximum level of education individuals achieved. Secondly, it contains information on occupation and non-occupation outcomes. Individuals with a high school degree, declaring to be in full time education are assumed to be enrolled in university.

The database contains detailed information about individual characteristics, employment status, individual income, household income and wealth, tenancy. Variables about labour market characteristics and income are spread across all the dataset: sector of activity, non-occupation outcomes (unemployment, retirement, in education), occupation status (blue collar, white collar, manager), number of worked hours, amount of fringe benefits, hours of overtime work. The structure of the survey allows to introduce these indicators both at an individual and at a household level. This is a desirable outcome, given the importance of family background characteristics in determining educational choices (Carneiro and Heckman [2002], Kane [1994], Petrongolo and San Segundo [2002] among others). It is worth noticing that the SHIW database has been used before to analyze research questions in the field of education economics (see for example Lucifora et al. [2000], Brunello and Miniaci [1999], Fiaschi and Gabriellini [2013]).

2.4.2 Sample selection

All main analyses are performed on a sub-sample of households where:

- one or more individuals live with their parents;
- at least one of them completed high school;
- individuals having completed high school are 18, 19 or 20 years old.[7]

The sample selection is motivated by regularities in the data that match with the findings of previous research. Firstly, individuals formally living with their parents (students and non-students) represent the large majority of young adults in the database. This evidence is consistent with the literature [Manacorda and Moretti, 2006] and with statistics on the topic. As the following graph shows, a relevant fraction of young Italian individuals aged between 18 and 34 live with their parents (see ISTAT [2014] for more data on the topic). Among them roughly half work, and a quarter are students.

[7] Further analyses (section 2.5.5) focus on the cyclicality of the enrollment decision at any year (including the first), and are performed on a subsample of individuals aged 18-25 years old.

Year	% of young, single individuals living with at least one of the parents	Employment status (a)				
		Employed	Looking for a job	Housewife	Student	Other Condition
			Males			
2001	68.0	52.4	18.6	-	25.2	3,7
2002	66.7	54.1	17.8	-	24.5	3,6
2003	66.2	53.6	16.4	-	25.8	4,2
2005	66.2	54.8	16.5	-	25.8	2,8
2006	67.3	53.3	18.7	-	25.9	2,1
2007	65.8	53.6	18.3	-	26.4	1,7
2008	66.2	51.3	19.3	-	28.3	1,2
2009	66.0	48.6	21.8	-	28.4	1,2
			Females			
2001	52,6	39.7	18.7	3.8	35.7	2,2
2002	52,6	39.3	19.1	3.4	36.3	1,9
2003	52,9	37.7	19.2	2.6	38.7	1,8
2005	52,5	38.6	16.6	3.2	39.7	1,9
2006	52,6	37.8	18.4	2.3	39.6	1,9
2007	52,4	39.6	18.1	2.1	38.2	2,0
2008	53,4	35.3	18.0	3.6	42.2	1,0
2009	50,9	34.1	20.5	3.6	40.2	1,6
			Males and Females			
2001	60,4	46.9	18.6	1.6	29.8	3,1
2002	59,7	47.6	18.4	1.5	29.7	2,9
2003	59,6	46.5	17.6	1.1	31.5	3,1
2005	59,5	47.7	16.5	1.4	31.9	2,4
2006	60,1	46.7	18.5	1.0	31.7	2,1
2007	59,1	47.4	18.2	0.9	31.6	1,8
2008	59,9	44.3	18.7	1.6	34.3	1,1
2009	58,6	42.5	21.3	1.5	33.4	1,4

(a) Per 100 Italian young, single individuals living with at least one of the parents

FIGURE 2.2: Percentage of 18-34 individuals living with their parents, disentagled by empolyment status (source:ISTAT)

Italian students' mobility within the country is relatively scarce. According to the Department for Education (Ministero dell'Istruzione, dell'Universita e della Ricerca, MIUR) during the academic year 2007/2008 80% of the students enrolled in the region in which they lived [MIUR, 2009]. Demarinis et al. [2014] analyses MIUR, and furnish detailed statistics on the topic that confirm the mentioned trend. The statistical evidence is consistent with academic contributions on the topic. Several scholars show indeed that Italian students' mobility is particularly low (Brunello and Cappellari [2008], Lucifora et al. [2000], Staffolani and Pigini [2012], Ordine and Lupi [2009]). More in detail, Ordine and Lupi [2009], confirm that Italian students tend to study in an institution located in their own region.[8] Given that students living with their parents represent the norm, this study mainly focuses on them.

Secondly, restricting to individuals living with their families allows to observe and control for household characteristics.[9] Kane [1994] studies the role of family background characteristics in explaining college enrollment for black students in the United States, and restricts to a similar subsample. He performs the analyses on households with dependent children aged 18 or 19. He considers the household head and their spouse as the young individuals' parents. This paper follows a comparable subsampling strategy. It restricts to households with dependent children, and identifies the household head and his spouse as the individual's parents. In tables and graph, ME stands for 'main earner' and identifies the household head. SE stands for 'second earner', and identifies a person married with the main earner and being the second earner in the household.

[8]This finding is generally attributed to the relative homogeneity of entry conditions, fees and quality of Italian universities [Brunello and Cappellari, 2008].

[9]The literature has long emphasized the role of family background characteristics in influencing educational choices. The list of contributions on this topic is potentially countless. Kane [1994] and Carneiro and Heckman [2002] are seminal examples that directly relate to the topic herein investigated.

The restriction to individuals aged 18 to 20 is justified by rather practical considerations. In Italy high school completion is (in general) the only requirement to enroll in university (Brunello and Cappellari [2008], Brunello and Miniaci [1999]). This happens, in most of the cases, during the year in which individuals turn 19. Many students do not continue their studies after high school, but a relevant fraction of them enroll in University straight after having graduated [Lucifora et al., 2000]. These represented 67.24% of the high school graduates in the academic year 2007/2008 [MIUR, 2009]. Enrollment takes place in September. Students who are born *after* September turn 19 by the end of the solar year and are still 19 when the academic year finishes. Students who are born *before* September turn 20 before the end of the academic year. As anticipated, the SHIW database is surveyed every two years. Restricting to individuals aged 18, 19 or 20 allows to include in the sample first year students who turn 20 during the first academic year. This choice has the shortcoming of including a small number of second-year students as well. This sacrifice is made not to compromise the sample size, which amounts to 4169 observations.[10]

2.4.3 Descriptive statistics

Table 2.2 contains descriptive statistics on the subsample considered.

Table 2.2: Summary statistics

	Observations	Mean	Std. Dev.	Min	Max
Number of Siblings	4169	1.16	.8664495	0	6
Age (ME)	4169	50.64068	5.605936	35	76
Educational Level (ME)	4169	3.400576	.960476	1	6
Labour Income (ME)	2505	21748.95	12414.8	430.5705	261276.7
Self Employment Income (ME)	856	25084.23	31844.5	216.3926	538213
Retirement (ME)	849	13114.01	8606.256	50	58237.4
Age (SE)	3746	47.93113	5.460728	31	70
Educational Level (SE)	3746	3.247731	.9686826	1	6
Labour Income (SE)	1498	17355.27	8228.799	288.5234	100602.2
Self Employment Income (SE)	395	18581.61	17570.98	216.3926	161463.9
Retirement (SE)	302	10899.04	6588.666	50	57318.32
HH Wealth	4169	308828.1	634538.7	-193504.5	2.45e+07
HH Financial Liabilities	4169	10314.57	38339.99	0	852017.9
HH Financial Assets	4169	31717.53	97663.03	0	2172125
HH Real Assets	4169	287425.2	595533.3	0	2.43e+07
Observations	4169				

Note. *Age* is measured in years. *Education* is a discrete ordered variable, bounded between 1 and 6, where 1 is primary school, and 6 is postgraduate education. Income and wealth variables are real annual indicators in , deflated using a CPI index, considered after taxes. *Number of Siblings* is measured in units. *Financial and Real Assets over Total wealth, Financial Exposition over Total Wealth* are shares, calculated as described in the variable name.

[10] The sample obviously does not contain first-year students who do not enroll straight after having completed high school, who took more than 5 years to complete high school, and who started primary school before 6.

Table 2.3: Mean Group Differences

Variable	Mean Difference	t-tstatistic
Female	-0.0611***	(-3.69)
Number of Siblings	0.0642*	(2.18)
Age (ME)	-0.459*	(-2.41)
Education (ME)	-0.608***	(-20.59)
Labour Income (ME)	-0.166***	(-7.79)
Slef Employment Income (ME)	-0.135*	(-2.26)
Transfers (ME)	-0.119	(-1.95)
Age (SE)	-0.335	(-1.73)
Education (SE)	-0.641***	(-20.98)
Labour Income (SE)	-0.214***	(-5.85)
Self Employment Income (SE)	-0.304**	(-2.76)
Transfers (SE)	-0.128	(-1.34)
HH Financial Liabilities	-3386.6**	(-3.11)
HH Financial Assets	-15612.4***	(-5.81)
HH Real Assets	-116241.4***	(-7.17)
N	4169	

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

The average individual with a high school degree, aged between 18 and 20 years old has 1.16 siblings, comes from the North in 37.84 % of the cases, from the South in 41,8 % of the cases, and from the Middle in 20.4 % of the cases (geographical information not reported in the table). The majority of their 'fathers' (as the main earner is male in 82.09 % of the cases) is employed and earns 21,750 euros per year (after taxes). Self-employed household heads play an important role in the sample, and earn more than the employed ones (25,064 euros on average). They are 50 years old, with a fairly low educational level. Their 'mothers' (as the second earner is female in 88.93 % of the cases) are slightly younger (almost 48 years old) and participate less in the labour market. If they work, they are mostly employed, with an average income of 17,355 euros. Their educational level is fairly low, and comparable to the one of the main earners.

Household wealth is on average substantial (slightly above 300,000 euros) and concentrated in housing (82%). Households in the sample are not particularly exposed from a financial point of view (10.4%). t-test results, contained in table 2.3, reveal that there are significant group-mean differences between enrolled and non-enrolled individuals. In particular, students tend to be females, to have less siblings, and to belong to a better socio-economic background than non-students. This preliminary result is consistent with the theory and with previous empirical evidence.

2.5 Empirical results

This section describes the empirical results. It starts by describing and motivating the techniques that will be implemented (2.5.1), and the specification chosen for the reduced form estimation (2.5.2). Section 2.5.3 displays and comments the main empirical results. Sections 2.5.4 to 2.5.7 provide a wide range of additional estimations and robustness checks to the findings obtained.

2.5.1 The estimation techniques

As anticipated, the paper makes use of binary and categorical choice models to analyze the cyclical behaviour of university enrollments.[11] Following Dellas and Sakellaris [2003], and in general the mentioned literature, the paper makes use of two alternative methodologies. The first is a two-step regression procedure. The enrollment decision is initially regressed on year dummies and on a set of controls. The year dummies coefficients are then regressed on a cyclical indicator (de-trended unemployment in the 15-24 class age; see section 2.3), to study how they correlate with the chosen variable. The following equations describe the two steps. Λ (from now onwards) is the logistic distribution. $Enolled_{it}$ (from now onwards) is a dummy variable, equal to one whether the individual is enrolled in tertiary education, equal to 0 otherwise. α is coefficient of interest, and describes the cyclical behaviour of the individuals in the sample.

$$Pr(Enolled_{it} = 1) = \Lambda(\gamma_0 + \sum_t \gamma_t YearDummy_{it} + controls + \zeta_{it}) \tag{2.1}$$

$$\gamma_t = \alpha_0 + \alpha_1 Indicator_t + \upsilon_t \tag{2.2}$$

The second technique is more straightforward. The decision of enrolling is regressed on the cyclical indicator (see above) and on the same set of controls. The coefficient β summarizes the cyclical behaviour of the individuals in the sample. The below equation describes the second technique.

$$Pr(Enolled_{it} = 1) = \Lambda(\beta_0 + Indicator_t\beta_1 + controls + \zeta_{it}) \tag{2.3}$$

[11] The choice is in line with most of the literature. Several early papers on the topic do not make use of individual data, and as a consequence drift from this tendency. Mattila [1982], for example, uses time series data. Pissarides [1981] postulates that the staying-on rate of 16 years old British individuals is a logistic function of a number of indicators, and applies OLS to a log-linearized transformation. Pissarides [1982] follows a similar rationale.

Further sections make use of alternative methodologies, and describe them in greater detail. It is worth noticing, however, that they all belong to the class of binary and categorical choice models.

2.5.2 The model specification

The specification proposed is the following:

$$Pr(Enolled_{it} = 1) = \Lambda(Var_t\gamma_0 + x_{it}\gamma_1 + x_{it}^{HH}\gamma_2 + x_{it}^{ME}\gamma_3 + x_{it}^{SE}{}_{\gamma 4} + x_i\gamma_5 + \zeta_{it}) \qquad (2.4)$$

where:

- Var_t is either year dummies or a cyclical indicator. The macroeconomic indicator chosen is unemployment levels in the 15-24 class age (thousand of individuals; see section 2.3). The variable is detrended using a Hodrick-Prescott filter ($\lambda = 100$). The unemployment rate is generally chosen for the same purpose in the cited literature (see section 2.2);
- vector x_{it} contains individual characteristics, such as gender, age and number of siblings;
- vector x_{it}^{HH} contains household level variables, such as financial liabilities, real wealth, financial wealth, tenancy;
- vectors x_{it}^{ME}, x_{it}^{SE} contain individual characteristics (age, gender, education) and labour market indicators attributable the parents of the individual (income, occupation status). Income (from labour, self employment or transfers) is the natural logarithms of real income after taxes, deflated using an OECD CPI index;
- vector z_i contains geographical indicators (area of origin, size of the town of origin, region of origin). This vector is intended to capture geographical or cultural differences in the enrollment decisions. Italy is characterized by deep north / south diversities. The northern part is more dynamic and overall richer, whilst the south is traditional and more agricultural. Italy also shows deep cultural differences between provinces and cities, captured by the town size dummies. The regional dummies are intended to control for the number of universities in the area (see Ordine and Lupi [2009]);

The present specification addresses a problem encountered in Dellas and Sakellaris [2003]. Their empirical model does not control for family background characteristics, which, as emphasized, are instead of great importance in explaining educational choices (see Kane [1994], Petrongolo and San Segundo [2002]).

2.5.3 Main empirical findings: first year enrollment decision

Tables 2.4-2.6 contain the main findings. Tables 2.4 and 2.5 detail about the first regression procedure. Table 2.6 displays result for the second estimation methodology (see section 2.5.1).

Table 2.4 shows the regression results when year dummies are introduced. The reported coefficients are marginal effects.[12] [13]

TABLE 2.4: Logit Models
Dependent Variable: Enrolled

	(1)	(2)	(3)	(4)	(5)
Year Dummies	x	x	x	x	x
Female	-	0.055***	0.069***	0.069***	0.065***
		(3.653)	(4.875)	(4.889)	(4.652)
Number of siblings	-	-0.013	-0.0042	-0.0053	-0.012
		(-1.429)	(-0.494)	(-0.599)	(-1.404)
Age (ME)	-	-	0.0041*	0.0042*	0.0023
			(1.775)	(1.781)	(1.144)
Age (SE)	-	-	-0.00021	-0.00019	-0.00023
			(-0.121)	(-0.179)	(-0.161)
Parental Education	-	-	x	x	x
Household Income	-	-	x	x	x
Household Wealth	-	-	-	x	x
Geographical Indicators	-	-	-	-	x
Tenancy	-	-	-	-	x
Observations	3,746	3,746	3,746	3,746	3,746

Note. z-statistics in parentheses. Clustered standard errors. Marginal Effects. *** p<0.01, ** p<0.05, * p<0.1

Females tend to enroll more. In the sample, being female increases the probability of enrolling by 6.5% (Table 2.4, column (5)). This finding is consistent with Staffolani and Pigini [2012].[14]

[12] For clarity of exposition, several coefficients are not displayed in the main text, but in appendix 1 (table 2.19). The present and the following sections make it clear when coefficients that are not reported in the tables are mentioned in the text.

[13] The sample chosen is close to being a repeated cross section, but technically it is not. Autocorrelation might represent a concern, given the (mildly) longitudinal structure of the database. It does not represent a major one, though. Firstly, very few individuals are observed more than once, and no one is observed more than twice. Secondly, inference is based on a clustered variance-covariance matrix. Thirdly, estimations of tables 2.4 to 2.6 have been performed on a cross-section of 19-20 years old individuals, and give qualitatively and quantitatively similar inference.

[14] Following Dellas and Sakellaris [2003], the paper also explores whether males and females show differences in their cyclical behaviour. They seem not to. Both girls and boys implement a procyclical behaviour. In the subsample of 18-20 years old individuals, when unemployment is above the trend by 10%, the probability if enrolling is reduced by 4% for females, by 3% for males (for males the result is not significant). In a subsample of 18-25 years old individuals (see section 2.5.5), when unemployment is above the trend by 10%, the probability if enrolling is reduced by 6% for females and by 4& for males. The results suggest that males and females have an overall comparable behaviour, with females being slightly more procyclical than males. Females might be more

Parental education plays an important role in explaining educational choices. When 'fathers' move upwards by an educational level, children are more likely to enroll in tertiary. As table 2.20 displays, when fathers step from high school to a bachelor's degree, the probability of enrolling increases by roughly 7 percentage points. When 'mothers' step from high school to a bachelor's degree, children are more likely to enroll by roughly 20 percentage points (table 2.19, column (5)).[15] Parental income, and in particular labour income, plays a role in explaining educational choices. When employed 'fathers' earn 10% more, children are more likely to enroll by 4 percentage points (Table 2.19, column (5)). When 'mothers' earn 10% more, children are more likely to enroll by 7 percentage points (Table 2.19, column (5)).

As table 2.19 shows, southern Italian students are more likely to enroll to University than northern ones. This finding is confirmed by external statistical results (see MIUR [2009]). Is this finding due to opportunity cost considerations? That is, are southern students more likely to enroll to university because of the lack of available options in the southern regions? Lucifora et al. [2000] report indeed that southern Italian regions face much more adverse economic conditions than in the North. Further evidence in needed to shed light on this result.

Table 2.21 in appendix 1 reports the main regression results, broken by geographical area (North, Middle, South). If Southern students were more likely to enroll to university because of opportunity cost considerations, the table would display a countercyclical enrollment decision for the 'south' subsample. Instead, as columns (2) and (3) suggest, enrollments are procyclical both in the Middle and in the South, and seem to be acyclical in the North. The reason for a higher propensity for enrollment in Southern regions does not seem to lie in opportunity cost considerations.[16] Although the evidence reported is not enough to draw definitive conclusions, Lucifora et al. [2000] observe that Southern Italian students enjoy higher returns to tertiary education than Northern ones. The paper endorses this view, in the belief that higher expected returns represent an incentive to enroll on their own, independently on cyclicality.

The findings described are consistent with the literature, and emphasize the role of family background characteristics in explaining educational choices (Petrongolo and San Segundo [2002], Kane [1994], among others).

Table 2.5 contains OLS regressions of year dummies coefficients on cyclical indicators.

risk averse in assessing the correlation influence of future economic conditions on current studying decisions, and henceforth might implement a slightly more procyclical behaviour than males. The results are displayed in appendix 1 (table 2.20).

[15] The paper refers to the main and the second earners respectively as 'fathers' and 'mothers' because most of the MEs are males and most of the SEs are females. See section 2.4.4.

[16] It might be argued that the enrollment decision for Southern students might be influenced by borrowing constraints. Brunello and Cappellari [2008] provide evidence that this is not the case.

TABLE 2.5: OLS regressions
Dependent Variable: Year Dummies Coefficients
Models of Table 2.4

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Unemployment (15-24)	-0.0188***	-0.0189***	-0.0200**	-0.0191**	-0.0203***	-0.0203**	-0.0034***
	(-4.753)	(-4.657)	(-3.304)	(-3.309)	(-3.969)	(-1.906)	(-3.64)
Constant	-0.0451**	-0.0452**	-0.0478*	-0.0456*	-0.0485**	-0.0485**	-0.0083***
	(-2.744)	(-2.706)	(-2.073)	(-2.099)	(-2.526)	(-1.942)	(-2.36)
Observations	9	9	9	9	9	9	9
R-squared	0.623	0.606	0.479	0.471	0.540	0.540	0.511

Note. Robust t-statistics in parentheses. Column (6) contains bootstrapped results (200 replications). Column (7) has been estimated using marginal effects. In columns (1)-(6), the unemployment coefficient represents the marginal change in the year dummy coefficient associated with unemployment being 1% above its trend. In column (7) the unemployment coefficient represents the marginal change in the probability of enrolling associated with unemployment being 1% above its trend. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Its main findings are also summarized by the following graph:

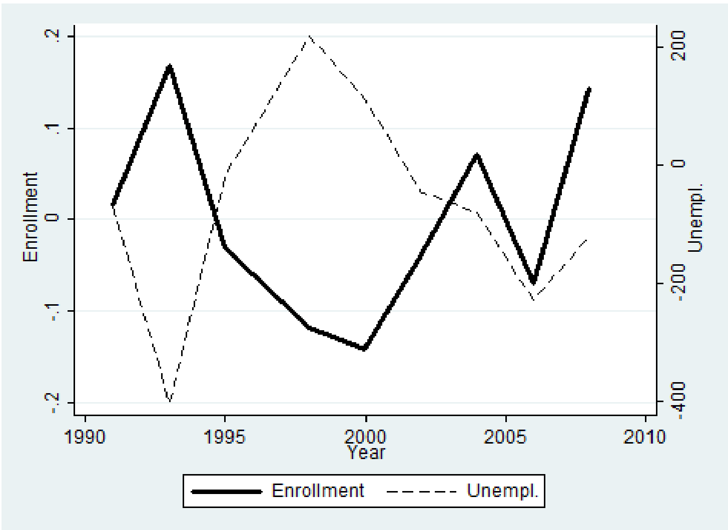

FIGURE 2.3: Year dummies coefficients on hp filtered unemployment

As the graph shows, the enrollment decision has a (roughly) negative correlation with unemployment. Consistently with the aggregate findings, the enrollment decision is characterized by a procyclical behaviour. As table 2.5 reports, when unemployment is above the cycle by 1%, the year dummy coefficients decrease by roughly 1.67 percentage points. The coefficient is significant at the 5% level, and the significance is robust to the introduction of an increasing number of regressors (table 2.5, columns (1)-(5)). Concerns might arise with respect to inference on the coefficients of table 2.5. These are indeed the results of a two-step estimation procedure. The standard errors of the second regression model need to be corrected using a bootstrapping procedure. Column (6) of table 2.5 contains the coefficient estimated in column (5), and standard errors taken from a bootstrapped variance covariance matrix.[17] The significance of the unemployment coefficient is robust to the resampling procedure.

The results of table 2.6 below confirm the evidence commented so far.

TABLE 2.6: Logit Models
Dependent Variable: Enrolled

	(1)	(2)	(3)	(4)	(5)
Unemployment 15-24 (cycle)	-0.0021	-0.0022	-0.0035*	-0.0036*	-0.0035*
	(-1.120)	(-1.090)	(-1.694)	(-1.661)	(-1.699)
Individual characteristics	-	x	x	x	x
Household income	-	-	x	x	x
Househol Wealth	-	-	-	x	x
Geographical Indicators	-	-	-	-	x
Marital Status	-	-	-	-	x
Tenancy	-	-	-	-	x
Observations	3,746	3,746	3,746	3,746	3,746

Note. z-statistics in parentheses. Clustered standard errors. Marginal effects. The estimator displayed represents the marginal change in the probability of enrolling associated with unemployment being 1% above its trend. *** p<0.01, ** p<0.05, * p<0.1

The enrollment decision is negatively correlated with unemployment. When unemployment is above the cycle by 10%, individuals are less likely to be enrolled by 3 percentage points (table 2.6, column (5)). The coefficient is significant at the 10% level, and robust to the introduction of an increasing number of regressors (table 2.6, columns (1)-(5)).[18]

[17] Estimation of the first-stage logit model has been replicated 200 times together with the OLS regression of the year dummy coefficients on the unemployment rate. Resampling took place at a cross-sectional unit level. See Kapetanios [2008] for a discussion of such bootstrapping procedure when both T and N are large. Kapetanios [2008] also justifies the adoption of such procedure when N is large but T is fixed, as in this case.

[18] As a robustness check, the regression model of table 2.6 has been estimated using unemployment rates in the 15-24 class age taken from alternative sources (OECD and ISTAT), rather than de-trended unemployment levels. Table 2.22 in appendix 1 displays the regression results. Consistently, the sign of the unemployment rate coefficients is negative. The coefficients, however, lose significance. This section argues that the main argument of the paper remains intact. De-trended levels are indeed more appropriate than unemployment rates to answer the research questions because they are informative on fluctuations around the trend. Unemployment rates do not contain this information.

The results of table 2.5 have been estimated using regression coefficients. This is certainly informative about the cyclical behaviour of enrollments. However, given the non-linear nature of logistic regression, the results of table 2.5 are hardly comparable to the results presented in table 2.6. To overcome this problem, the two-step procedure that lead to the estimation of table 2.5 has been replicated using marginal effects, rather than regression coefficients. The result is contained in table 2.5, column (7). The reported coefficient would represent the marginal change in the probability of enrolling associated with a 1% point increase in unemployment. Comparison between table 2.5 column (7) and table 2.6 column (5) immediately reveal how the two magnitudes do not differ from each other, as expected.

To summarize, this section shows that the first year enrollment decision is characterized by a procyclical pattern. The result is robust to the implementation of two different techniques and to the introduction of an increasing number of regressors. This baseline result is in general at odds with the empirical evidence, and in particular with opportunity cost considerations. It is however consistent with the presence of borrowing constraints [Christian, 2007]. It also might reflect the influence of expected future averse labour market conditions on the enrollment decision [Micklewright et al., 1990].

2.5.4 First year enrollment decision and labour market outcomes.

As mentioned above, the enrollment and the working decision are intrinsically substitutes [Perli and Sakellaris, 1998]. An individual deciding to study indeed usually renounces to be in full time employment. More in general, the enrollment decision is a substitute to alternative occupational and non-occupational choices (employment, unemployment and non-participation. See Petrongolo and San Segundo [2002], Dellas and Sakellaris [2003]).

Working activities and human capital accumulation are not perfect substitutes, though. Individuals often supply a small amount of labour to finance their education. The section exploits this (imperfect) substitutability to estimate a bivariate probit model, where the dependent variables are the individual enrollment and working decision. The technique relies on the restrictive but necessary assumption of joint normality of the error terms. The purpose of this analysis is to jointly estimate two reduced form estimations and to compare and contrast their results. The joint estimation is not performed with identification purposes, but to improve the efficiency of the reduced form coefficients (see Chiappori and Salanie [2000]).

Table 2.7 provides estimations results of bivariate probit models. Column (1) contains estimation results when year dummies are introduced in the specification. Column (2) contains estimation results when unemployment is on the right hand side. For the two specifications, the null hypothesis of no correlation between the error terms ($\rho = 0$) is strongly rejected. The test result supports the choice of estimating the models jointly.

TABLE 2.7: Bivariate Probit Regression. Dependent Variables: Enrolled, Employed

	(1)		(2)	
Dependent Variable:	Enrolled	Employed	Enrolled	Employed
Year Dummies	x	x	-	-
Unemployment (15-24)	-	-	-0.003*	0.002***
			(-1.855)	(2.802)
Female	0.072***	-0.012*	0.073***	-0.014**
	(4.641)	(-1.835)	(4.697)	(-2.102)
Number of siblings	-0.014	0.006	-0.011	
	(-1.414)	(1.496)	(-1.126)	(1.241)
Age (ME)	0.0041	0.00023	0.0038	0.0019
	(1.536)	(0.519)	(1.311)	(0.797)
Education (ME)	0.084***	-0.026***	0.082***	-0.025***
	(7.685)	(-5.870)	(7.506)	(-5.523)
Age (SE)	0.00023	-0.0027*	-0.00015	-0.0026
	(0.010)	(-1.670)	(-0.039)	(-1.618)
Education (SE)	0.094***	-0.023***	0.086***	-0.019***
	(8.317)	(-4.787)	(7.719)	(-3.951)
Household Income	x	x	x	x
Household Wealth	x	x	x	x
x				
Geographical Indicators	x	x	x	x
Tenancy	x	x	x	x
Observations	3746	3746	3746	3746
Wald Test on ρ (p-value)	0.000		0.000	

Note. Robust z-statistics in parentheses. Marginal effects. *** p<0.01, ** p<0.05, * p<0.1

The results displayed in table 2.7 show a clear symmetry. When a regressor is significant for one decision, it generally shows a significant correlation of the opposite sign for the other one. The magnitudes do not show exact symmetry, and are in general milder for the employment decision. For example, if 'father's' education increases by one level the probability of being enrolled increases by 3.5 percentage points. The probability of working is instead reduced by 0.5 percentage points only (table 2.7, column (2)).

The enrollment decision shows a procyclical pattern, as column (2) suggests. The coefficient of the cyclical component is similar to the ones of table 2.6. The working decision shows instead a countercyclical pattern. The result is summarized by the following graph:

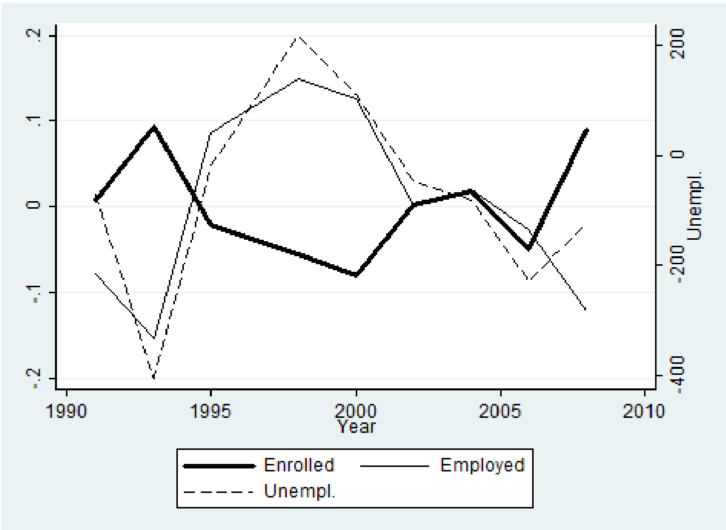

FIGURE 2.4: Year dummies coefficients on hp filtered variables

This evidence is not trivial. Although this section does not claim identification, the finding that young individuals depending on their family supply labour countercyclically is overall suggestive. Indeed, if enrollments were procyclical due to household borrowing constraints *only*, nothing would preclude the working decision to be procyclical as well (as it should be). It doesn't happen to be the case, suggesting that the reason for a procyclical enrolling choice might lie somewhere else.

Table 2.8 displays OLS estimations of the year dummies coefficients of table 2.7, column (1) on the cyclical indicator. The findings are consistent with the ones of table 2.7 and with the ones of the previous section, and do not need further comments.

Table 2.8: OLS estimations
Dependent Variable: Year Dummies Coefficients
Model of table 7, column (1)

	Enrolled	Employed
Unemployment (15-24)	-0.00976**	0.0204***
	(-3.210)	(6.377)
Constant	-0.0233*	0.0488**
	(-2.107)	(2.919)
Observations	9	9
R-squared	0.477	0.681

Note. Robust t-statistics in parentheses. Marginal effects reported. The number displayed represents the marginal change in the year dummy coefficient associated with unemployment being 1% above its trend. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Petrongolo and San Segundo [2002] and Dellas and Sakellaris [2003] explore the issue of substitutability among labour market activities by estimating multinomial logistic regressions. They consider being out of the labour force, and/or unemployment, as additional choices to employment and education. A similar estimation has been performed, where the dependent variable is a categorical indicator equal to 1 if the individual is a student, equal to 2 if the individual works, equal to 3 if the individual is unemployed, equal to 4 if the individual is out of the labour market. The results of this estimation are consistent with the general results, and are displayed in appendix 1 (table 2.23). When being enrolled is the omitted category, unemployment increases the odds of working (odds ratio: 1.026*, t-statistic: 1.889), of being unemployed (odds ratio: 1.007, t-statistic: 0.674) and of being out of the labour market (odds ratio: 1.068***, t-statistic: 2.905), with respect to the ones of studying. The results show that bad economic conditions make human capital accumulation less appealing than all the other competing decisions, and are henceforth consistent with the evidence commented so far.[19]

2.5.5 The enrollment decision at any year.

The graph and the regression table contained in section 2.3 show that aggregate first year enrollments in tertiary education follow a procyclical pattern. As mentioned briefly, this result is robust to the introduction of different enrollment indicators; among them, the total number of students in tertiary education. This section explores whether the individual behaviour is consistent with this finding, and analyses the enrollment decision at any year (including the first).

[19] The multinomial logit model depends on the so-called 'independence of irrelevant alternatives' (IIA) assumption. This implies that the odds ratios of two alternatives (say being in education rather than being working) is not affected by the presence of other alternatives (say, unemployment). Despite this assumption is rather restrictive, neither Dellas and Sakellaris [2003] nor Petrongolo and San Segundo [2002] mention this problem. The IIA assumption can be tested (Hausman-McFadden test). Some studies showed however that the test performs poorly, even in large samples [Cheng and Long, 2007].

2.5.5.1 Pooled logistic regressions

In the Italian university system the first year enrollment decision does not strongly determine one's choices in the following years. Firstly, the drop-out rate in Italy has traditionally been particularly high (Becker S. [2001], Lucifora et al. [2000], Brunello and Miniaci [1999]).[20] Secondly, an individual can take more than the minimum required period to graduate. They actually do, as Becker S. [2001] reports. The present section exploits the panel structure of the database to analyze the cyclical behaviour of the enrollment choice at any year. To limit the noise, it concentrates on students 'on time'.[21]

The estimations proposed in the previous sections are now performed on a sub-sample of individuals aged 18 to 25. This age restriction is made to minimize the amount of lagged behind students. Indeed, if an individual aged 18-19 takes 3-4 years to complete a bachelor's (or one-tier) degree, they would finish when they are 23-24. They would terminate when they are 25, if the degree takes five years to complete.[22]

The models commented in this section are pooled logistic regressions. Assuming that the error term is not correlated with the independent variables, pooled logistic regression coefficients are consistent under non particularly restrictive assumptions. In the regression tables, inference is based on a clustered variance-covariance matrix to take non-spherical disturbances into account. The assumption that the error terms are not correlated with the regressors is not likely to be met, though. Pooled logistic regressions only work as a baseline result. Section 2.5.5.3 follows Méndez and Sepúlveda [2012] in dealing with this issue.

Table 2.9 contains the regression results. Panel 1 displays the binary choice models. Column (1) introduces year dummies in the specification, column (2) introduces unemployment. The year dummies coefficients are then regressed on unemployment. Results are shown in panel 2. The specification now contains a second order polynomial in age, to control for nonlinear influences of aging on the enrollment decision (see Dellas and Sakellaris [2003]).

[20]Brunello and Miniaci [1999] mention that less than one third of the students who start university eventually graduate. The reference is dated, but relevant for the present contribution, as data refer to the 1991-2010 time span.

[21]Regression results on the sample of lagged-behind students (the so-called 'fuori corso') shows that they are not sensitive to the business cycle. The results are displayed in appendix 1 (table 2.24).

[22]As Lucifora et al. [2000] and Becker S. [2001] describe, bachelor's degree were introduced into the Italian university system with a 2001 reform. Before that date, degrees were organized as one-tier programs, and last either 4 or 5 years, depending on the subject.

Table 2.9: Pooled logistic regression.
Dependent Variable: Enrolled. Individuals aged 18-25.

Panel 1: Main Regressions	(1)	(2)
Year Dummies	x	-
Unemployment (15-24)	-	-0.0043***
		(-4.716)
Female	0.077***	0.078***
	(7.831)	(7.927)
Number of siblings	-0.010	-0.0089
	(-1.631)	(-1.267)
Age	-0.287***	-0.293***
	(-6.647)	(-6.793)
Age (squared)	0.0055***	0.0053***
	(5.420)	(5.547)
Age (ME)	0.0041***	0.0042***
	(3.231)	(3.265)
Education (ME)	0.078***	0.077***
	(12.021)	(11.865)
Age (SE)	0.0017	0.0015
	(0.692)	(0.551)
Education (SE)	0.096***	0.091***
	(14.441)	(13.736)
Household Income	x	x
Household Wealth	x	x
Geographical Indicators	x	x
Tenancy	x	x
Observations	11437	11437
Panel 2: OLS regression of year dummies coefficients on unemployment		
Unemployment (15-24)	-0.0256***	-
	(-3.664)	-
Constant	-0.0612	-
	(-1.845)	-
Observations	9	-
R-squared	0.553	-

Note. Panel 1: z-statistics in parentheses. Clustered standard errors. Marginal effects. Panel 2: Robust t-statistics in parentheses. The number displayed represent the maginal change in the year dummy coefficient associated with unemployment being 1% above its trend. *** p<0.01, ** p<0.05, * p<0.1

All regressors confirm the results obtained so far. Parental education, household income and household wealth positively influence the enrollment decision (income and wealth are not reported). Females enroll more. 'Father's' age is now significant: when the ME is 10 year older, the probability of enrolling goes up by 4 percentage points. The enrollment decision still shows a procyclical pattern, as the following graph, together with panel 1, column (2), and panel 2 confirm. The magnitude of the cyclical influence is fairly similar, although slightly higher than the ones of the previous tables. When unemployment is above the cycle by 10%, the individuals are less likely to enroll by 4 percentage points (compared to 3 percentage points in table 2.6, column (5) See table 2.9, column (2)).[23] The following graph provides a representation of the results:

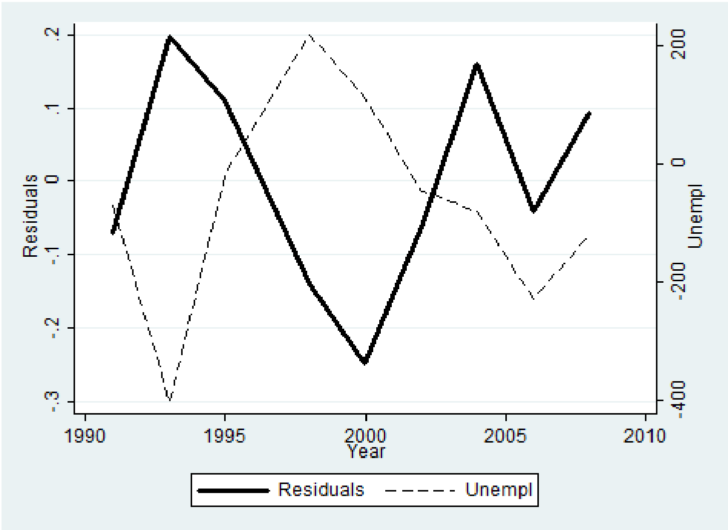

FIGURE 2.5: Year dummies coefficients on hp filtered variables

2.5.5.2 Enrollment decision (at any year) and labour market outcomes

Following section 2.5.4, the present section discusses the relationship between the enrollment decision and labour market outcomes for the 18-25 years old subsample. It briefly comments on estimation results of a bivariate probit model and a multinomial logit model (with the same dependent variable of section 2.5.4). To avoid redundancies, the regression tables are displayed in appendix 1 (tables 2.26 and 2.27). The bivariate probit model results are qualitatively similar to the ones of table 2.7, although with different magnitudes. Females tend to enroll more (marginal effect: 0.105^{***}, t-statistic: 10.782) and work less (marginal effect: -0.044^{***}, t-statistic: -5.800) than males. Parental education stimulates the enrollment choice, and discourages the working decision (the marginal effects for both parents are, roughly: 0.10 for the enrollment decision, -0.05 for the working decision. The four coefficients are significant at the 1% level). The cyclical pattern of the two choices is the same as before. Unemployment is negatively correlated with the enrollment decision (marginal effect: -0.006^{***}, t-statistic: -5.228) and positively correlated with the employment one (marginal effect: 0.005^{***}, t-statistic: 5.477).

[23]Similarly to what discussed at the end of section 2.5.3, as a robustness check the main regression has been performed using young unemployment rates. The results, displayed in table 2.25 in appendix 1, are consistent with the ones of table 2.22. The same argument developed in footnote 18 applies here.

The multinomial logit results are comparable to the ones commented in section 2.5.4. In a recession period, tertiary education seems to be the less appealing option among different occupational and non-occupational outcomes even in the 18-25 years old subsample. When being enrolled is the omitted category, unemployment increases the odds of working (odds ratio: 1.029***, t-statistic: 4.902), of being unemployed (odds ratio: 1.015**, t-statistic:2.445) and of being out of the labour market (odds ratio: 1.089***, t-statistic: 6.335), with respect to the ones of studying.

2.5.5.3 Fixed-effects

The omission of variables correlated with the regressors leads to bias in the coefficients. This is a standard result, and an often unavoidable price when non-experimental data are used. Longitudinal data, however, allow to limit (but not completely solve) this problem. By using individuals as their own controls, fixed effect models control for time invariant unobservable individual heterogeneity [Allison, 2009]. Assuming that several time invariant characteristics are correlated with the regressors, estimating a fixed effect model should reduce (but not neutralize) the omitted variable bias.

Méndez and Sepúlveda [2012] provide a similar discussion in their contribution. As the authors notice, the binary-choice analogue of a linear fixed effect regression is conditional logit (Andersen [1970], Chamberlain [1980]).[24] Conditional logistic regression has the appealing benefit of being computationally efficient. It has however the drawback of depending on the conditional independence assumption, which is quite restrictive and often violated. Moreover, Kwak and Wooldridge [2009] show that the conditional logit estimator is particularly sensitive to violations of conditional independence. Despite its appealing properties and widespread use, conditional logit estimation may create more problems than the ones it solves.

A popular alternative to conditional logit, robust to violations of conditional independence, is Correlated Random-Effects Probit (CRE) [Chamberlain, 1980]. The principle of this approach is simple. The conditional distribution of unobservable heterogeneity is modeled as a function of a set of covariates, and substituted in the model specification. The model is eventually estimated using a random-effect model probit regression [Wooldridge, 2011]. The CRE approach overcomes several of the limitations of conditional logit: it allows the identification of marginal effects; it

[24] Although conditional logit is often used to control for (individual) fixed effects, conditional logistic regression has a wider range of applications. Conditional logistic regression estimates a binary choice model by conditioning on all observed choices *within a stratum*. In this way it can also be informative about how individuals choose among a range of possible alternatives (the alternatives being the strata themselves). Terry Long [2004] uses conditional logit to study the determinants of individual college decision. The author explains very well how conditional logit works in her paper (pages 277-278). For this kind of applications, conditional logit depends on the IIA assumption (Terry Long [2004], Staffolani and Pigini [2012], Gibbons and Vignoles [2012]). Staffolani and Pigini [2012], among others, proposes the use of nested logit to overcome this problem.

keeps regressors that do not show within-group volatility; it keeps cross sectional units that do not show volatility in the dependent variable.[25] CRE has the drawback, however, of modeling individual heterogeneity as an arbitrary function of the regressors [Méndez and Sepúlveda, 2012].

Following Méndez and Sepúlveda [2012], table 2.10 contains estimation results for a conditional logit and a correlated random-effect probit model.

TABLE 2.10: Random Effect and Conditional Logistic Regressions. Dependent Variables: Enrolled. Individuals aged 18-25.

	(1) Conditional Logit	(2) CRE Probit
Unemployment (15-24)	-0.036***	-0.020***
	(-2.953)	(-3.608)
Number of siblings	0.099	0.259**
	(0.666)	(2.435)
Labour Income (ME)	0.081***	0.031*
	(3.284)	(1.918)
Self Eployment Income (ME)	0.061**	0.025
	(1.985)	(1.276)
Transfers (ME)	-0.010	0.023
	(-0.412)	(1.360)
Labour Income squared (SE)	0.026	0.021
	(1.021)	(1.277)
Self Eployment Income (SE)	0.064*	0.042*
	(1.758)	(1.929)
Transfers (SE)	-0.038	0.019
	(-1.144)	(0.931)
Household Financial Wealth	0.000072	0.000075
	(0.234)	(0.343)
Household Financial Liabilities	-0.000051	0.000052
	(-0.636)	(0.541)
Household Real Wealth	-0.000048	-0.000047*
	(-1.500)	(-1.731)
Female	-	0.455***
		(7.539)
Age	-	-0.318***
		(-14.739)
Geographical Indicators	-	x
Parental Education, age and gender	-	x
Within-group means of regressors	-	x
Constant	-	1.794***
		(3.167)
σ_u^2		1.295***
		(14.087)
Observations	1734	11437

Note. Robust z-statistics in parentheses. The numbers displayed are estimation coefficients. *** p<0.01, ** p<0.05, * p<0.1

[25] On a survey on the pros and cons of fixed effects logit, see Allison [2009].

Some clarifications are needed before commenting the regression results. In the conditional logit model, the specification proposed is less rich than the ones of the previous tables for three different reasons. Firstly, conditional logit estimation cannot be performed using regressors that do not show any within-group volatility (as gender, gender of the parents, geographical area). Secondly, regressors that show little volatility within individuals, but significant variations across individuals inflate the conditional logit standard errors [Allison, 2009]. As a consequence, some indicators (such as parental education) have been discarded *a priori*. Thirdly, the specification does not contain the age polynomial. Given the short time span, the introduction of the age polynomial resulted fairly redundant, and captured the most of the regression significance, leading to misleading results.[26]

In the Correlated Random Effect Probit model, unobservable heterogeneity is modeled as function of the means of all time variant regressors that show a certain degree of within-individual volatility (in fact, the regressors of the conditional logit specification). This follows Wooldridge [2011].

Table 2.10 reports the estimation coefficients, as in conditional logit marginal effects cannot be computed. Interestingly, the unemployment measure preserves a negative and significant coefficient in both regression models. The enrollment decision robustly shows a procyclical pattern, and the result holds even after controlling for individual, time-invariant unobservable heterogeneity. It is however desirable to compare the results of table 2.10 with the ones displayed in the above sections. In column (2), when unemployment is above the cycle by 10%, the probability of being enrolled is reduced by 4% (significant at the 1% level, t-statistic: -4.769). This magnitude is overall similar to the one of table 2.9, column (2). Pooled probit regression (not reported) leads to the same result. Neglected (time invariant) individual heterogeneity does not seem to represent a major concern in the estimated models.

2.5.6 Reconciling the findings with the theory

This section discusses the results of additional estimations. It investigates whether the presence of a procyclical behaviour in the enrollment decision is a robust and consistent result. Moreover, it tries to reconcile the empirical evidence with theoretical predictions.

[26]The regression results of conditional logit containing the age polynomial are displayed in appendix 1 (table 2.28).

2.5.6.1 Household borrowing constraints

The literature generally attributes a procyclical enrollment decision to borrowing constraints [Christian, 2007]. The specifications proposed somehow take borrowing constraints into account. Indeed, they introduce household income and wealth on the right-hand-side, in this way controlling for collateral to obtain credit. Nevertheless, the influence of borrowing constraints can be tested more explicitly. The paper tests for the influence of household borrowing constraints in three different ways. It introduces among the regressors an interaction term between a wealth indicator and a cyclical one, and checks the sign and the significance of its coefficient (see Méndez and Sepúlveda [2012]). It performs the regressions on a subsample of household that do not show any financial liabilities, to check whether their cyclical behaviour differs from the overall one. Finally, iy follows Christian [2007] in restricting the analysis to a subsample of home owners, again to see whether their cyclical behaviour is different from the overall one.

Table 2.11 contains the regression results. All regressions are performed using the specification of table 2.6, column (5). Actually, table 2.11, column (1) displays the same regression results of table 2.6, column (5), reported again to facilitate the comparison with the other models. The table also reports a test of whether the unemployment coefficient of column (1) is different from the one in the corresponding column. The rationale of the comparison is the following: if borrowing constraints were *the only reason* for a procyclical enrollment choice, the regression results should show a countercyclical behaviour when borrowing constraints are not present.

TABLE 2.11: Logistic Regression. Dependent Variables: Enrolled.

	Panel 1: Individuals aged 18-20			
	(1) Baseline	(2) With Interaction	(3) No Financial Liabilities	(4) Home Owners
Unemployment (15-24)	-0.0035*	-0.0041*	-0.0038*	-0.0029
	(-1.699)	(-1.931)	(-1.781)	(-1.112)
Controls	x	x	x	x
Observations	3746	3746	2482	2909
Test (p-values)	-	0.66	0.51	0.46

	Panel 2: Individuals aged 18-25			
	(1) Baseline	(2) With Interaction	(3) No Financial Liabilities	(4) Home Owners
Unemployment (15-24)	-0.0047***	-0.0048***	-0.0052***	-0.0049***
	(-4.716)	(-3.427)	(-4.346)	(-3.514)
Controls	x	x	x	x
Observations	11437	11437	7712	8979
Test (p-values)	-	0.13	0.24	0.16

Note. z-statistics in parentheses. Clustered standard errors. Marginal effects. The unemployment coefficient represents the marginal change in the probability of enrolling associated with unemployment being 1% above its trend. *** p<0.01, ** p<0.05, * p<0.1

The analysis proposed by Méndez and Sepúlveda [2012] (panels 1 and 2, column (2)) does not allow to conclude that borrowing constraints influence the enrollment decision. The unemployment coefficient shows the same, or even stronger procyclicality than in column (1). In both panels, the two coefficients do not show statistical difference from each other. The interaction term has a magnitude close to zero, and it is not significantly different from zero (not reported). The method proposed by Méndez and Sepúlveda [2012] does not allow to detect any influence of household borrowing constraints on the enrollment decision.

Column (3) contains regression results performed on a subsample of households with no financial liabilities. In both panels, the unemployment rate has a significant and negative coefficient. The chi-squared test does not allow to distinguish the coefficients of column (3) from the ones of column (1), suggesting that the magnitude of this correlation is not influenced by household borrowing constraints.

The analysis proposed by Christian [2007] is the only one showing that borrowing constraints have some influence on the enrollment decision, and only in panel 1. In the subsample of home-owners, the magnitude of the unemployment coefficient is lower than in column (1), and becomes statistically insignificant (table 2.11, panel 1, column 4). Panel 2, however, does not show a consistent evidence.

To summarize, the unconstrained samples exhibit, at most, a *milder* procyclicality. This suggests that *household* borrowing constraints do not have a strong influence on the cyclical movement of Italian enrollments.

The paper follows Christian [2007] in assessing whether the cyclical behaviour of the enrollment decision varies across income groups. The results are displayed in table 2.12.

Separate regressions are performed for individuals living in households at different points of the income distribution. The evidence found is comparable to the one of table 2.11. Individuals aged 18-20 and living in households above the third of the income distribution show a mild and non significant countercyclical behaviour (table 2.12, panel 1, column (4)). The cyclicality of first year enrollments seems to be mainly driven by households placed between the median and the third quartile the income distribution (table 2.12, panel 1, column (3)). Enrollments at any year show instead a prcyclical behaviour across all points of the income distribution (table 2.12, panel 2). No income group exhibits a strong countercyclical behaviour, suggesting once again that household borrowing constraints only mildly influence the enrollment decision.

Table 2.12: Logistic Regression. Dependent Variables: Enrolled. Individuals aged 18-20.

	Panel 1: Individuals aged 18-20			
	(1)	(2)	(3)	(4)
	$Income < \alpha(25)$	$\alpha(25) < Income < \alpha(50)$	$\alpha(50) < Income < \alpha(75)$	$Income > \alpha(75)$
Unemployment (15-24)	-0.0051	-0.0037	-0.0063*	0.0018
	(-1.335)	(-0.988)	(-1.771)	(0.386)
Controls	x	x	x	x
Observations	851	895	968	995
	Panel 2: Individuals aged 18-25			
	(1)	(2)	(3)	(4)
	$Income < \alpha(25)$	$\alpha(25) < Income < \alpha(50)$	$\alpha(50) < Income < \alpha(75)$	$Income > \alpha(75)$
Unemployment (15-24)	-0.034***	-0.011	-0.044***	-0.020**
	(-3.434)	(-1.205)	(-4.731)	(-2.116)
Controls	x	x	x	x
Observations	2508	2786	2971	3164

Note. z-statistics in parentheses. Clustered standard errors. Marginal effects. The unemployment coefficient represents the maginal change in the probability of enrolling associated with unemployment being 1% above its trend. *** p<0.01, ** p<0.05, * p<0.1

To further investigate whether borrowing constraints influence the enrollment decision, this section also concentrates on the employment status of the main and the second earner. More in detail, it displays regression results estimated on subsamples of households where:

- the main earner is unemployed;
- either the main or the second earner is unemployed;
- the main earner is unemployed and the second earner stays at home (housewife or staying-at-home dad)

under the assumption that household in which the main earners are in the above employment statuses are more likely to be borrowing constrained. The following table contains the regression results:

Table 2.13: Logistic Regression. Dependent Variables: Enrolled.

	18-25 subsample			18-20 subsample		
	(1)	(2)	(3)	(4)	(5)	(6)
	ME unempl.	ME or SE unempl.	ME unempl., SE at home	ME unempl.	ME or SE unempl.	ME unempl., SE at home
Unemployment (15-24)	0.0013	-0.0033	-0.0055	0.00067	0.002	0.0038
	(0.137)	(-0.635)	(-0.610)	(0.002)	(0.299)	(0.239)
Controls	x	x	x	x	x	x
Observations	322	489	136	97	152	41

Note. z-statistics in parentheses. Clustered standard errors. Marginal effects. The unemployment coefficient represents the maginal change in the probability of enrolling associated with unemployment being 1% above its trend. *** p<0.01, ** p<0.05, * p<0.1

The above analysis confirms that borrowing constraints do not have a major influence on the enrollment decision of young individuals. All coefficients (apart from the regression in column (3)) show a lower magnitude than in the main regression, or even a sign reversal. Moreover, in all regressions the unemployment rate coefficient loses significance. The t-statistics correspond to very high p-values in all columns (see for example column (4)). These regression results provide additional evidence that the procyclicality of the enrollment decision cannot be attributed to borrowing constraints.

2.5.6.2 Future labour market conditions

Borrowing constraints are not the only possible reason for a procyclical enrollment decision. Micklewright et al. [1990] suggest that, during a recession, individuals might demand for less education if they expect future economic conditions to be negatively influenced by the current ones. This result holds if the unemployment rate is homogeneous across educational levels. If individuals with a better education are less likely to be influenced by adverse economic conditions, the demand for education can actually be stimulated by a recession [Tumino and Taylor, 2015].

To test the above statements, this section introduces both young and adult unemployment rate on the right-hand-side, and checks the sign and the significance of the two indicators (following Petrongolo and San Segundo [2002] and Tumino and Taylor [2015]). Young unemployment rate captures the substitution effect between the studying and the working decision. According to opportunity cost considerations, it should be characterized by a negative and significant coefficient. Adult unemployment rate instead approximates future economic conditions. It should be characterized by a negative and significant coefficient, if future economic conditions are expected to be correlated with current ones *and* the unemployment rate is homogeneous across educational levels. It should be characterized by a positive and significant coefficient, if individuals insure themselves against adverse future economic conditions by acquiring more education [Tumino and Taylor, 2015].

Table 2.13 contains the estimation results. Column (1) contains results for the 18-20 years old subsample. Column (2) contains results for the 18-25 years old subsample. The set of controls is the one of table 5, column (5).

The findings of Tumino and Taylor [2015] and Petrongolo and San Segundo [2002] are not confirmed by the ones herein reported. Firstly, the enrollment decision reacts negatively to the young unemployment rate even when controlling for future economic conditions. Secondly, adult unemployment rate is positively correlated with the enrollment choice. This result is pasrticularly strong in magnitude and highly significant for the 18-25 years old individuals (table 2.13, column (4)). Following the interpretation of Micklewright et al. [1990] and Tumino and Taylor [2015], individuals seem to demand for more education to insure themselves against adverse future economic conditions.

Table 2.14: Logistic Regression. Dependent Variables.

	(1) 18-20 subsample	(2) 18-25 subsample
Unemployment (15-24)	-0.0033	-0.0091***
	(-0.934)	(-4.339)
Unemployment (25-64)	0.00055	0.0046**
	(0.003)	(2.035)
Controls	x	x
Observations	3746	11437

Note. z-statistics in parentheses. Clustered standard errors. Marginal effects. The unemployment coefficients represent the marginal change in the probability of enrolling associated with unemployment being 1% above its trend. *** p<0.01, ** p<0.05, * p<0.1

2.5.6.3 Reverse causality

Reverse causality between the dependent variable and the main regressor (de-trended young unemployment levels) represents a possible source of concern for the models estimated above. The underlying reasoning is trivial: one more student in university represents on young unemployed individual less. As a consequence, the negative correlation found in the above regression might be attributed to this very simple fact, rather than to cyclicality.

The below table contains regression results aiming to addressing this issue. De-trended young unemployment is now substituted with 5 different indicators:

- de-trended unemployment levels in the 25-64 class age
- de-trended total unemployment levels
- unemployment rate in the 25-64 class age
- total unemployment rate
- employment to population ratio in the 15-24 class age

The reasons underlying the introduction of these particular indicators are the following. With respect to the first indicator (de-trended levels of unemployment in the 25-64 class age), the variable should not be or should be only mildly affected by the enrolling decision. Young individuals are indeed simply not in the count of 25-64 unemployment levels. The same reasoning applies to the second indicator (de-trended levels of total unemployment), which should be only mildly affected by the enrollment decision. Unemployment rates (the third and the fourth indicator) should not be affected by the enrollment decision at all: although one more student represents one person less among unemployed individuals, it also represents one person less in the labour force. As a consequence, the computation of unemployment rates should not be influenced by the enrolling decision. Finally, the employment to population ratio should be in principle negatively correlated with the enrollment decision: one more student, indeed, represents one employed person less, but not one person less in the population. As a consequence, if this regressor is found to be positively correlated with the enrollment decision, the result would hint towards a robust procyclicality.

TABLE 2.15: Logistic Regression. Dependent Variables: Enrolled.

	Panel 1: Individuals aged 18-20				
	(1)	(2)	(3)	(4)	(5)
Total Unempl. (de-trended)	-0.0046*				
	(-1.917)				
Unempl. 25-64 (de-trended)		-0.0034*			
		(-1.941)			
Total unempl rate			-0.0048		
			(-0.527)		
Unempl. rate 25-64				-0.0029	
				(-0.319)	
Empl-pop ratio 15-24					0.013***
					(3.357)
Controls	x	x	x	x	x
Observations	3746	3746	3746	3746	3746
	Panel 2: Individuals aged 18-25				
	(1)	(2)	(3)	(4)	(5)
Total Unempl. (de-trended)	-0.0078***				
	(-5.365)				
Unempl. 25-64 (de-trended)		-0.0045***			
		(-3.946)			
Total unempl rate			-0.0082*		
			(-1.798)		
Unempl. rate 25-64				-0.0078	
				(-1.373)	
Empl-pop ratio 15-24					0.011***
					(4.206)
Controls	x	x	x	x	x
Observations	11437	11437	11437	11437	11437

Note. z-statistics in parentheses. Clustered standard errors. Marginal effects. The unemployment coefficient represents the marginal change in the probability of enrolling associated with unemployment being 1% above its trend. *** p<0.01, ** p<0.05, * p<0.1

The results suggest that the bias due to reverse causality does not represent a major concern for the regression results displayed so far. All signs are consistent with procyclicality, and most of the indicators are significant. Overall, magnitudes are comparable to the main results (table 2.6), especially for the de-trended unemployment levels in the 25-64 class age. The coefficients of total unemployment (both in levels and in rates) are slightly higher, suggesting that a the coefficients displayed in table 2.6 might be biased downwards. Nevertheless, the substance of the main argument (procyclicality of enrolments) does not seem to be affected by the reverse causality issue.

2.5.6.4 Supply side considerations

The regression results displayed so far concentrate on the demand side of the enrollment decision. The models describe indeed the individual decision on enrolling, conditional on personal characteristics and the business cycle. It is however worth considering that cyclical behaviour of university enrollments might be influenced by supply side considerations. Universities might indeed decide the number of available places on the basis of economic fluctuations. If in a recession (booming) period universities cut (increase) their budgets and reduce (raise) the number of available places, the procyclical behaviour observed might be attributed to the lack (availability) of available places during a recession (booming) period. The regression models partially take this into account, though. Given that 80% of the Italian university students study in their region of origin (see section 2.4.2.), regional dummy control indeed for the supply of education for a more than relevant fraction of the sample. There's however a remaining 20% of students who base their choice on the supply of university places located somewhere else in the country. It is clear that for these individuals the regional dummies do not represent a good control for the supply of university places. This paragraph argues however that supply-side constraints do not represent a major problem in the Italian university system. Two reasons are invoked in support of this statement.

The first reason is that limited access universities do not represent the norm in the Italian university system. As Staffolani and Pigini [2012] observe, only some faculties systematically admit a limited number of student on the basis of an entry test (Medicine and Architecture). Moreover, as ISTAT [2002] reports, the introduction of constraints to the supply of university places in these faculties (the report mention Architecture in particular) is relatively recent. Although the practice of limiting the number of students had been present from the eighties, the first law regulating this issue nationwide dates back to 1999, with effects on the academic year 2000/2001.

Given the Italian regulation, only a fraction of the enrollments can be attributed to limited access places. According to 'Il Sole 24 Ore' (the main Italian financial and economic newspaper), during the academic year 2008/2009 121,404 first year university students were enrolled in limited-access courses. This number represents 36% of the total number of first year enrollments (as reported by ISTAT).

The second reason is that Italian universities are characterized by an excess of available capacity. The following map, taken from Demarinis et al. [2014], shows the number of enrolled students over the number of available places in the academic year 2008/2009, broken by province. Only 4 out of 110 Italian provinces (coloured in green) are characterized by an excess of demand.

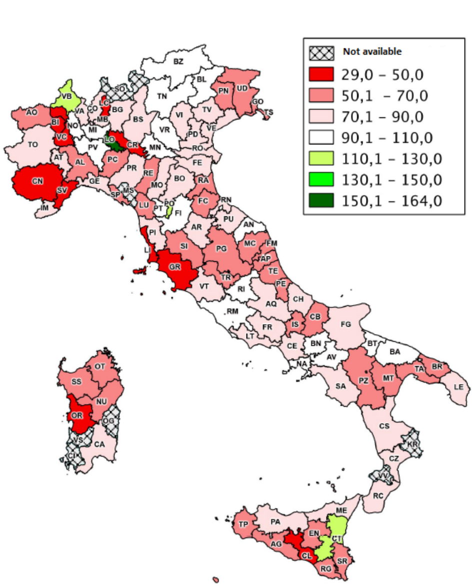

FIGURE 2.6: Percentage of first year enrollments over available places

The evidence reported in this section points towards non-binding supply-side constraints in the period considered. Firstly, as a matter of fact, Italian universities do not exploit their full capacity. Secondly, only a fraction of the enrollments can be attributed to limited access courses, and these were legally introduced only half-way through the time span herein considered. Thirdly, students always have the option of enrolling to a non-limited access faculty if they fail an entry exam. As a consequence, it can be concluded that supply-constraints are not binding in the Italian university system for the period considered.

2.5.6.5 Individuals living on their own

Before proposing an interpretation for the evidence found, it might be useful to summarize what displayed so far:

1. aggregate enrollments show a procyclical pattern;
2. the individual enrollment decision is consistent with the aggregate pattern. This finding is robust to different specifications and techniques;
3. borrowing constraints do not seem to play a role in shaping the enrollment decision. The procyclicality of enrollments cannot be explained using this interpretation;
4. although individuals seem to insure themselves against future adverse economic conditions by acquiring more education, they respond procyclically to current economic conditions.

This paper proposes a straightforward interpretation to solve the puzzle emerged in the data. It argues that borrowing constraints operate at an individual level. The Italian educational system does not rely on unsecured loans for the financing of tertiary education. It is also characterized by very limited assistance to households with financial need [Lucifora et al., 2000]. Young individuals heavily rely on family resources to fund their human capital acquisition. The paper suggests that during booming periods, when household resources are abundant, parents are eager to finance the continuation of their children' studies. But during recession moments, when family resources are scarce, young individuals are not given funds to enroll to university. They might even be asked to contribute to the family resources. This explanation would be consistent with the countercyclical working decision.

To support this interpretation table 2.14 contains estimation results of some of the models proposed above on a subsample of 18-25 years old individuals *living on their own*.[27] Supposedly, individuals on their own do not (totally) depend on family resources to make their decisions.

[27] The sample selection restricts to household heads or spouses of the household head. Standard errors are clustered by household.

Moreover, they are more likely to have collateral, and as a consequence an easier access to credit than individuals living with their parents.

Table 2.16: Binary Choice Models. Dependent Variables: Enrolled. Individuals aged 18-25.

	(1) Pooled Logit	(2) CRE Probit	(3) Financially Exposed (1)	(4) Financially Exposed (2)
Unemployment (15-24)	0.0056*	0.079*	-0.025***	-0.021***
	(1.849)	(1.741)	(-3.089)	(-2.680)
Controls	x	x	x	x
Within-group Mean of Regressors	-	x	-	-
Observations	697	701	110	99

Note. z-statistics in parentheses. Clustered standard errors. Marginal effects. The unemployed coefficient represents the marginal change in the probability of enrolling associated with unemployment being 1% above its trend. *** p<0.01, ** p<0.05, * p<0.1

As expected, in this subsample the enrollment decision shows a clear countercyclical pattern (table 2.14, column (1)). A 10% increase in the unemployment rate increases the probability of enrolling by 5 percentage points. The result is robust to the estimation of a correlated random effects model (table 2.14, column (2)). Columns (3) and (4) restrict the sample to individuals whose financial liabilities are respectively 10% (column (3)) and 15% of their gross wealth. These individuals are assumed to be borrowing constrained. The sign reversal shows that in this subsample borrowing constraints play a role. Borrowing constrained individuals living on their own implement a procyclical enrollment decision.

To summarize, and conclude, the enrollment decision of individuals living on their own has a cyclical behaviour that is consistent with opportunity costs and borrowing constraints considerations, and matches with most of the empirical evidence. The results suggest that young individuals living with their families, who generally do not have access to credit, simply rely on family resources (and as a consequence, on their fluctuations) to finance their education.

2.6 Conclusions

The present paper investigates whether Italian tertiary education enrollments are influenced by the business cycle. As a general result, it finds that the enrollment decision is procyclical. This pattern is present in aggregate data and is confirmed by individual analyses.

This result is found on a sample of young individuals living with, or dependent on, their families. It holds for individuals enrolling for the first time (18-20 years old individuals) and for individuals enrolling at any year (18-25 years old individuals). It is robust to estimation with different techniques, and to the introduction of an increasing number of controls.

The relationship between the results and theoretical predictions is not immediately evident. Firstly, in the same subsample the working decision has a countercyclical behaviour. Secondly, neither household borrowing constraints, nor the relationship between current and expected economic conditions are supported in the data as possible explanations to the procyclical pattern.

An interpretation that would make all the empirical findings consistent with each other is that borrowing constraints operate at an individual level. Individuals living with their families generally do not have access to credit. As a consequence, they would implement a procyclical enrollment decision because they are borrowing constrained themselves, and heavily dependent on household resources. This would explain why they implement a procyclical enrollment decision; why they implement a countercyclical working decision (as they contribute to the family reource during bad periods); why enrollments have a procyclical pattern even if households are not borrowing constrained.

Further analyses confirm the validity of this explanation. The same empirical models, estimated on a sample of 18-25 years old individuals living on their own, show that these individuals implement a countercyclical enrollment decision if not borrowing constrained, and a procyclical one if borrowing constrained. This finding is consistent with most of the empirical evidence, and with theoretical predictions.

Theoretical contributions show that countercyclical human capital accumulation would have desirable long-run economic consequences. As a policy implication, then, individuals should be released from their borrowing constraints. They should be given more conspicuous financial support, an easier access to credit and an easier access to the labour market during their university years.

Appendix 1: additional results

TABLE 2.17: OLS regressions
Dependent Variable: First Year Enrollments in Tertiary Education
1977 - 2010

	(1)	(2)	(3)	(4)	(5)	(6)
Employment ISTAT	0.868**	-	-	-	-	-
	(2.573)					
Self Employment (ISTAT)	-	1.418***	-	-	-	-
		(3.900)				
GDP (ISTAT)	-	-	0.443**	-	-	-
			(2.186)			
Unempl. ISTAT	-	-	-	-1.839***	-	-
				(-5.143)		
Unempl. ISTAT (15-24)	-	-	-	-	-1.357***	-
					(-5.143)	
GDP (OECD)	-	-	-	-	-	0.474**
						(2.143)
Constant	2.647	3.099**	2.227	2.616**	2.267	2.034
	(1.450)	(2.212)	(1.080)	(2.185)	(1.572)	(0.966)
Observations	21	21	20	21	21	21
R-squared	0.250	0.348	0.199	0.586	0.509	0.233

Note. z-statistics in parentheses. Bootstrapped standard errors (100 replications). The displayed coefficients are elasticities. *** p<0.01, ** p<0.05, * p<0.1

TABLE 2.18: OLS regressions
Dependent Variable: Enrollments rates in Tertiary Education
1989 - 2010

	(1)	(2)	(3)	(4)	(5)	(6)
Empl-pop ratio	1.517***					
	(5.711)					
Empl-pop ratio 25-64		2.585***				
		(4.301)				
Total unempl rate			-2.614***			
			(-3.090)			
Unempl rate 25-64				-2.293***		
				(-9.005)		
Empl-pop ratio 15-24					-2.492***	
					(-6.230)	
Unempl rate 15-24						-1.070***
						(-9.305)
Constant	-63.435***	-85.264***	47.420***	50.966***	96.361***	58.750***
	(-3.935)	(-3.220)	(7.682)	(20.100)	(8.838)	(17.915)
Observations	11	11	11	11	11	11
R-squared	0.632	0.493	0.335	0.810	0.671	0.820

TABLE 2.19: Logistic Regression. Dependent Variable: Enrolled.

	(1)	(2)	(3)	(4)	(5)
Year Dummies	x	x	x	x	x
Female		0.055***	0.069***	0.069***	0.067***
		(3.653)	(4.875)	(4.889)	(4.759)
Number of siblings		-0.013	-0.004	-0.005	-0.012
		(-1.429)	(-0.494)	(-0.599)	(-1.408)
Age (ME)			0.004*	0.004*	0.003
			(1.775)	(1.781)	(1.264)
ME:Postgraduate			0.474***	0.454***	0.434**
			(2.710)	(2.583)	(2.473)
ME:Degree			0.404***	0.392***	0.393***
			(4.372)	(4.246)	(4.270)
ME:High School			0.328***	0.320***	0.326***
			(3.772)	(3.683)	(3.773)
ME:Lower Secondary			0.240***	0.236***	0.243***
			(2.782)	(2.747)	(2.838)
ME:Primary			0.171**	0.167*	0.173**
			(1.980)	(1.944)	(2.025)
Age (SE)			-0.000	-0.000	-0.000
			(-0.121)	(-0.179)	(-0.075)
SE:Postgraduate (collinear)			0.000	0.000	0.000
			(.)	(.)	(.)
SE:Degree			0.402***	0.393***	0.385***
			(5.313)	(5.162)	(5.003)
SE:High School			0.193***	0.187***	0.185***
			(3.086)	(2.983)	(2.862)
SE:Lower Secondary			0.076	0.072	0.077
			(1.243)	(1.174)	(1.230)
SE:Primary			0.024	0.022	0.022
			(0.398)	(0.359)	(0.353)
Labour Income (ME)			0.003	0.004	0.004*
			(1.337)	(1.433)	(1.660)
Labour Income (SE)			0.006***	0.006***	0.007***
			(3.364)	(3.299)	(3.698)
Transfers (ME)			0.0035	0.0036	0.0041
			(0.998)	(1.011)	(1.406)
Transfers (SE)			0.0041	0.0039	0.0042
			(1.144)	(1.093)	(1.367)
Self Eployment Income (ME)			0.0043	0.0037	0.0044
			(1.551)	(1.080)	(1.522)
Self Eployment Income (SE)			0.005*	0.004	0.005*
			(1.945)	(1.570)	(1.761)
Gender			-0.025	-0.022	-0.020
			(-0.847)	(-0.741)	(-0.688)
Household Financial Wealth				0.000073	0.000074
				(0.222)	(0.675)
Household Financial Liabilities				-0.000065	-0.000068
				(0.917)	(1.161)
Household Real Wealth				0.000054	0.000055
				(1.490)	(0.946)
Home owner					0.064**
					(2.109)
Tenant					-0.012
					(-0.354)
With right of redemption					-0.080
					(-0.916)
Use without charge (collinear)					0.000
					(.)
North					-0.035*
					(-1.747)
South					0.032
					(1.535)
Observations	3746	3746	3746	3746	3746

Marginal effects; t statistics in parentheses

(d) for discrete change of dummy variable from 0 to 1

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.20: Logistic Regression. Dependent Variable: Enrolled.

	18-25		18-20	
	M	F	M	F
	(1)	(2)	(3)	(4)
Unempl. (15-24)	-0.0063***	-0.0059***	-0.0047*	-0.0031
	(-4.161)	(-3.562)	(-1.731)	(-1.549)
Controls	x	x	x	x
Observations	5753	5684	1791	1955

Note. z-statistics in parentheses. Clustered standard errors. Marginal Effects. *** p<0.01, ** p<0.05, * p<0.1

Table 2.21: Logistic Regression. Dependent Variables: Enrolled.

	North	Middle	South
	(1)	(2)	(3)
Unempl. ISTAT (15-24)	-0.00011	-0.0066*	-0.0059**
	(-0.129)	(-1.705)	(-2.182)
Controls	x	x	x
Observations	1384	720	1640 height

Marginal effects; t statistics in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.22: Logistic Regression. Dependent Variables: Enrolled.

	(1)	(2)
Unempl. rate 15-24 (ISTAT)	-0.0014	-
	(-0.543)	
Unempl. rate 15-24 (OECD)	-	-0.0026
		(-0.950)
Control	x	x
Observations	3746	3746

Marginal effects; t statistics in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.23: Multinomial Logit. Omitted category: enrolled. 18-25 class age.

	Employed	Unemployed	Out of LF
	(1)	(2)	(3)
Unempl. (15-24)	1.026*	1.0079	1.068***
	(1.889)	(0.674)	(2.905)
Controls	x	x	x
Observations		3746	

Note. z-statistics in parentheses. Odds Ratios. *** p<0.01, ** p<0.05, * p<0.1

Table 2.24: Logistic Regression. Dependent Variables: Enrolled. Lagged behind students ('fuori corso')

Unemployment (15-24)	0.00044
	(0.045)
Controls	x
Observations	471

Marginal effects; t statistics in parentheses

(d) for discrete change of dummy variable from 0 to 1

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.25: Logistic Regression. Dependent Variables: Enrolled.

	(1)	(2)
Unempl. rate 15-24 (ISTAT)	-0.0014	
	(-0.690)	
Unempl. rate 15-24 (OECD)		-0.0021*
		(-1.772)
Observations	11437	11437

Marginal effects; t statistics in parentheses

(d) for discrete change of dummy variable from 0 to 1

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE 2.26: Bivariate Probit Regression. Dependent Variables: Enrolled, Employed

	Enrolled (1)	Employed (2)
enrolled		
Unemployment (15-24)	-0.0064***	0.0059***
	(-5.228)	(5.477)
Female (d)	0.105***	-0.044***
	(10.782)	(-5.800)
Number of siblings	-0.008	0.006
	(-1.279)	(1.220)
Age (ME)	0.001	0.001
	(0.793)	(0.962)
Education (ME)	0.090***	-0.059***
	(13.144)	(-10.795)
Age (SE)	-0.004***	0.002*
	(-2.914)	(1.766)
Education (SE)	0.106***	-0.047***
	(15.128)	(-8.510)
Labour Income (ME)	0.007***	-0.001
	(4.316)	(-0.699)
Labour Income (SE)	0.008***	-0.003***
	(6.169)	(-3.323)
Pension (ME)	0.006***	0.000
	(3.109)	(0.120)
Pension (SE)	0.005***	-0.002
	(2.679)	(-1.461)
Self Empl. Income (ME)	0.009***	-0.006***
	(4.889)	(-4.369)
Self Empl. Income (SE)	0.007***	-0.005***
	(3.802)	(-3.173)
Female (ME)	-0.004	0.031*
	(-0.215)	(1.886)
Financial Assets	0.000*	-0.000
	(1.700)	(-0.987)
Financial Liabilities	-0.000***	0.000***
	(-4.097)	(3.982)
Housing Wealth	0.000***	-0.000**
	(4.466)	(-2.184)
Geographical Indicators	x	x
Observations	11437	11437

Marginal effects; t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

TABLE 2.27: Multinomial Logit. Omitted category: enrolled. 18-25 class age.

	Employed (1)	Unemployed (2)	Out of LF (3)
Unempl. (15-24)	1.029***	1.015**	1.089***
	(4.902)	(2.445)	(6.335)
Controls	x	x	x
Observations		11437	

Note. z-statistics in parentheses. Odds Ratios. *** p<0.01, ** p<0.05, * p<0.1

TABLE 2.28: Conditional Logistic Regression. Dependent Variables: Enrolled. Individuals aged 18-25.

Unemployment (15-24)	-0.010
	(-0.640)
Number of siblings	0.184
	(1.131)
Labour Income (ME)	0.025
	(0.863)
Transfers (ME)	-0.005
	(-0.179)
Self Eployment Income (ME)	0.055
	(1.485)
Labour Income squared (SE)	0.021
	(0.719)
Self Eployment Income (SE)	0.084*
	(1.906)
Transfers (SE)	0.010
	(0.257)
Household Financial Wealth	-0.000071
	(-0.202)
Household Financial Liabilities	0.000065
	(0.003)
Household Real Wealth	-0.000059
	(-1.140)
Age	-2.414***
	(-3.949)
Age squared	0.043***
	(3.086)
Observations	1734

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Chapter 3

Labour Income Risk Hedging in Italian Households.

Summary

This paper empirically assesses the relationship between idiosyncratic labour income volatility and investment in risky assets, at a household level. The study is performed on a representative sample of Italian households (SHIW, Bank of Italy). Reduced form estimations show that idiosyncratic labour income volatility crowds out household allocation in risky assets. The result is robust to the introduction of an increasing number of regressors. Inference based on a bootstrapped variance covariance matrix strengthens this main result. Estimating the model using different dependent variables and techniques does not influence the overall findings. Additional estimations address the endogeneity of labour income volatility and show that the negative relationship found is compatible with the presence of hedging motives in household behaviour.

3.1 Introduction

As opposed to theoretical predictions [Merton, 1971], most households choose not to invest in risky assets showing positive expected returns (the so-called stock-holding puzzle: Haliassos and Bertaut [1995]). Moreover, if they do, they generally dedicate a small fraction of their wealth to them [King and Leape, 1998]. Calibration of theoretical models show exactly the opposite result [Kocherlakota, 1996]. The literature has been trying to explain this phenomenon in several different ways. A motivation often considered is uninsurable labour income risk. In the real world, incomplete financial markets prevent households from insuring their idiosyncratic earnings shocks. This background risk in turn discourages both participation in the risky assets market and risky asset allocation (Guiso et al. [1996],Campbell [2006]). In support of this statement, several studies show that a positive correlation between labour income shocks and stock market returns reduces households' stock market allocation (Bonaparte et al. [2014], Arrondel et al. [2010]).

This paper assesses whether household labour income risk negatively influences investment in risky assets showing positive expected returns. Moreover, it explores whether this relationship can be attributed to a hedging behaviour. To address this question, the paper implements reduced form estimations, making use of Italian longitudinal data (SHIW).[1] It proposes the construction of a novel measure of idiosyncratic risk. The measure is in turn introduced on the right-hand-side of regression models, in which the dependent variable is the share of risky assets over financial wealth. The estimations allow to conclude that labour income volatility crowds out investment in risky assets, and that the evidence found is compatible with a hedging behaviour.

The paper contributes to the literature in several ways. It adds to the labour income risk hedging literature, and provides additional evidence that idiosyncratic labour income volatility negatively influences household allocation in risky assets. The result is robust to an increasing number of controls, and to hypothesis testing based on a bootstrapped variance-covariance matrix.

The coefficient of labour income volatility is likely to be affected by omitted variable bias. To conclude that labour income risk has a direct influence on portfolio decisions, the endogeneity problem needs to be addressed. In the spirit of Betermier et al. [2012], the paper contains several robustness checks aimed at showing that human capital risk has an actual influence on investment decisions. The findings are also generally consistent with the presence of hedging motives in household investment behaviour. Although the paper does not solve the endogeneity problem, it adds to the literature by emphasizing a direct connection between uninsurable human capital risk and portfolio choices (similarly, in the rationale, to Betermier et al. [2012]).

An important contribution of the paper is the creation of a novel measure of idiosyncratic labour income volatility. The literature commonly follows two approaches to calculate labour income risk measures. The first is to calculate earnings volatility using the deviations between a self-assessed measure of income and its actual realization (see for example, Arrondel and

[1]Several scholars approached research questions on the topic using structural equation modelling. See Haliassos and Michaelides [2003]); Sanroman [2015]; Alan [2006]; Alan [2012] ; Bonaparte et al. [2012]; Perraudin and Sorensen [2000] ; Gomes and Michaelides [2005]. Reduced form estimation has been chosen as it allows to study the influence of labour income volatility on investment in risky assets, at the same time controlling for a wide number of indicators that cannot be as flexibly introduced in a structural model.

Calvo-Pardo [2014], Arrondel et al. [2010]; see Guiso et al. [1996], using the SHIW database). The second is to postulate a 'canonical' earnings model, and to calculate idiosyncratic income volatility as the standard deviation of labour earnings' growth rate for each individual (Heaton and Lucas [1997], Betermier et al. [2012], Bonaparte et al. [2014]). Both approaches present shortcomings that the measure herein proposed tries to overcome. Section 3.4 details more about this.

The final contribution of the paper is methodological. Given the nature of the dependent variable (mixed distribution with a mass point at zero, long tail), similar empirical studies implement censored data models (Tobit) or a sample selection models (Heckit). The paper argues that these techniques are not fully appropriate. Santos Silva et al. [2014] recently proposed a technique that better fits data with the characteristics displayed in this research. This paper implements their methodology. Moreover, it provides evidence that this choice is preferable with respect to the implementation of sample selection models. Sections 3.5.1 and 3.5.2 provide an accurate discussion on this topic.

The remainder of the paper is structured as follows: section 3.2 contains a literature review. Section 3.3 describes the database chosen and the sub-sample selection. Section 3.4 describes the construction of the labour income risk score. Section 3.5 describes the estimation process and comments on the estimation results. Section 3.6 concludes.

3.2 Literature review

The benchmark model generally used as a starting point for household portfolio theory describes a utility maximizing agent who lives for one period and is endowed an amount invested in a portfolio of assets. The portfolio is liquidated at the end of the period to finance consumption. Markets are complete. Under this assumption, modeling labour income would be redundant. Individuals can indeed transfer their labour income risk to the financial market. The agent has to choose between a risk-free asset, with returns R, and a risky asset, with a positive return premium on the risk-free asset $\tilde{R}$. A standard consumption optimization problem (in an Arrow-Debreu economy) would predict that the optimal share of wealth (w) invested in the risky asset is approximately equal to:

$$w \approx R\frac{E[\tilde{R}]}{\tilde{\sigma^2}}\frac{1}{\theta} \tag{3.1}$$

where $E[\tilde{R}]$ and $\tilde{\sigma^2}$ are respectively the expected return premium and the expected volatility of the risky asset, and θ is the individual relative risk aversion coefficient [Gollier, 2002, p. 35]. It is worth noticing that the result does not not admit a corner solution.

Calibrations of the above model on U.S. data show that households with an average coefficient of risk aversion should invest a relevant fraction of their wealth in risky assets (Merton [1971]; Heaton and Lucas [1997]). Kocherlakota [1996] shows that a coefficient of risk aversion equal to 4 should correspond to a share of wealth invested in risky assets equal to 55%. Empirical evidence does not support these results. Firstly, households do not invest as much as predicted in risky assets. To explain the numbers observed in the U.S., individuals should be characterized by disproportionately large coefficients of risk aversion (around 40). This result is known as the stock-holding puzzle (Haliassos and Bertaut [1995]), and is the micro equivalent of the equity premium puzzle [Miniaci and Weber, 2002, p. 145]. Moreover, most households actually do not invest at all in risky assets. This result is at odds with the predicted absence of a corner solution.

Attitude towards risk alone cannot explain the stock market participation puzzle. The literature proposes several reasons for this problem. Perraudin and Sorensen [2000] argue that the presence of monitoring costs prevents households from holding all possible kinds of assets. Sanroman [2015] focuses her attention on the existence of participation costs in the asset market. Alan [2006] focuses her attention on disastrous events. Alan [2012] concentrates instead on participation costs. So does Vissing-Jorgensen [2002], who also introduces transaction costs in her model.

Several contributions explain the stock-holding puzzle through non-diversifiable labour income risk. When markets are not complete and labour income risk in non-tradable, a higher labour income volatility induces households to invest in safer assets to diversify their uncertainty. Chang et al. [2013] stress the role of non-diversifiable labour income risk in explaining the stock-holding puzzle. Arrondel and Calvo-Pardo [2014] emphasize labour income uncertainty *and* borrowing constraints as possible causes of the corner solution problem. Knupfer et al. [2013] use an exogenous unpredictable event to show that labour income shocks have a strong negative impact on household portfolio allocation in risky assets.

A number of contributions, including the present one, empirically investigate the determinants of household allocation in risky assets and study the presence of hedging motives in household finance decisions. Bonaparte et al. [2014] and Arrondel et al. [2010] show that a positive income-risky assets returns correlation induces households to invest less in risky assets, consistently with the presence of hedging motives. Betermier et al. [2012] show that households adjust their portfolios when household members switch their job, consistently with the hedging motive. Massa and Simonov [2006] provide opposite evidence with respect to Bonaparte et al. [2014], arguing that households may invest more in assets the returns of which are more correlated with their labour income because of familiarity reasons.

To construct a measure of idiosyncratic income risk, this paper analyses the longitudinal covariance structure of houehold earnings. The literature on the topic dates back to the seventies. Lillard and Willis [1978], Lillard and Weiss [1979], MaCurdy [1982], Abowd and Card [1989] represent seminal contributions. In the cited studies the authors pool together micro data and analyse the covariance structure of earnings assuming homogeneity of income processes across individuals. Baker [1997] deals with the problem of individual heterogeneity, comparing the "profile heterogeneity model", in which the earnings profile is individual specific, with a unit root model that does not allow any individual heterogeneity. The author concludes that the former specification better fits the data.

The debate on what model to use is still open. Browning et al. [2010] decompose labour income, letting each individual in the sample have their own income process (complete individual heterogeneity). Meghir and Pistaferri [2004], on the contrary, construct a more complicated specification, without allowing individual heterogeneity. Recent work by Altonji et al. [2013] remarkably contributed to the literature on the topic by developing a multivariate analysis of individual earnings. The cited seminal contribution by Abowd and Card [1989] already moved towards this direction. Altonji et al. [2013] develop a rich structural specification describing the behaviour of wages and earnings in the light of: labour market transitions, experience accumulation, worked hours changes. The authors combine the literature on earnings decomposition with the search-and-matching one.[2]

3.3 Database description and sample selection

The data are taken from the Survey of Household Income and Wealth of the Bank of Italy (SHIW from now on).[3] The SHIW database is a random sample of nearly 8,000 households per year. Although it is available from 1977 it shows an actual panel structure only from 1989 onwards. For this reason, the analyses proposed pertain to the 1989 - 2010 time span, and make use of eleven waves in total.

[2] See also Guvenen and Smith [2010] for an analysis of the volatility of earnings in the light of consumption decisions.

[3] The SHIW dataset has already been used in the literature to study household portfolio choices, albeit with different techniques (Sanroman [2015]; Pelizzon and Weber [2009]; Guiso and Jappelli [2002]; Guiso et al. [1996]).

The dataset contains information about individual characteristics, employment status, individual income, household income and wealth, tenancy. In particular, the information about household financial wealth is particularly rich. Variables about labour market characteristics and incomes are spread across all the dataset. The database has information about the sector of activity, non-occupation outcomes (unemployment, retirement, in education), occupation status (blue collar, white collar, manager), number of worked hours, amount of fringe benefits, hours of overtime work. Information about individual education and human capital is unfortunately relatively limited.

All analyses of this paper are performed on a sub-sample of households where:

- the main earner (ME in equations and tables) is employed;
- the main earner is married;
- the person married with the main earner is the second earner in the household and is employed as well (the "second earner", SE in equations and tables);
- the main and second earner are observed more than 5 times.

In the sub-sample so chosen 97% of household income comes from labour income. This has two desirable consequences. Firstly, the savings allocation is essentially in the hands of the two main earners. Secondly, household income volatility is almost fully attributable to labour income volatility. This reduces the noise of the estimations, and enhances the statistical precision of the analyses performed. The sample size amounts to 327 households, for a total number of 2244 observations. According to the specifications proposed, the sample size is generally reduced for collinearity reasons.

3.4 Labour income risk score

This section proposes and develops a novel measure of idiosyncratic labour income risk. Section 3.4.1 frames the procedure in the literature. Section 3.4.2 discusses the creation of the labour income risk indicator.

3.4.1 Theoretical background

Income risk can be trivially seen as the difference between expected income and its actual representation. Following this logic and canonical theoretical results, several authors compute earnings volatility using self-assessed survey measures of future earnings. Arrondel et al. [2010] and Arrondel and Calvo-Pardo [2014] follow this approach. Guiso et al. [1996] is of particular interest for this paper. The authors indeed make use of the SHIW database to construct a measure of this kind. As these very same contributions mention, this procedure presents several drawbacks [Guiso et al., 1996, pages 162-163]. This paper does not support their approach for two reasons. Firstly, not all survey databases contain information regarding subjective expectations on earnings. Secondly, a measure of this kind neglects the actual earning generating process of the individuals.[4]

A second possibility is to make use of available data to build an income volatility measure. As Browning et al. [2010] emphasize, the canonical representation of income processes allows to express the growth rate of earnings as the sum of permanent and transitory shocks.[5] Under this assumption, several papers use the standard deviation of labour income growth rates as a measure of labour income risk (Betermier et al. [2012], Bonaparte et al. [2014]).[6]

This approach has the desirable feature of being feasible with any income longitudinal data. It has, however, two shortcomings. Firstly, the risk measure so calculated neglects the influence of deterministic components of labour income. The literature agrees in estimating earning processes *after* having cleansed income from deterministic growth factors. Research often concludes that the unit root model is a possible or optimal representation of the data. This happens, however, after regressing income on year dummies and potential experience (see Abowd and Card [1989], Baker [1997]). Secondly, such measure completely neglects idiosyncratic heterogeneity in the earnings process, and postulates a unit root representation for all individuals. It is however debated whether the canonical model represents the best fit for the data (see Baker [1997], Browning et al. [2010]).

Following the literature on labour income dynamics, this paper aims at creating a labour income risk measure that overcomes the mentioned drawbacks. The following section details more about the creation of this measure.

[4]It can be argued that household members with rational expectations know their income generating process. However, as Guiso et al. [1996] and Arrondel et al. [2010] emphasize, the quality of individuals' assessment depends on the wording of the question. In addition, this does not represent the only source of noise in individuals' answers (see Meyer [2015] for a recent contribution on the topic). Even if household members were aware of their income process, their self-assessment would be far from being precise.

[5]The 'consensus' model (as Browning et al. [2010] call it) defines income as the sum of a permanent and transitory component, and the permanent component as a unit root process. Theoretical papers generally use this specification: see Gomes and Michaelides [2005], Alan [2006], Alan [2012], among others.

[6]Such a measure clearly does not disentangle permanent from transitory risk. For an analysis of how permanent and transitory components of labour income risk affect investment choices, see Koo [1999], Gomes and Michaelides [2005].

3.4.2 The creation of a measure of idiosyncratic risk

As stated above, one of the problems of the labour income volatility measures commonly found in the literature is that earnings contain deterministic components. Following a standard practice in the literature on labour income dynamics, a pooled regression of (log) labour income on year dummies and age dummies is initially performed (see Abowd and Card [1989]; Baker [1997]; Meghir and Pistaferri [2004]; Borella [2001]; Browning et al. [2010]).[7] The year dummies control for 'aggregate risk' (as defined in Meghir and Pistaferri [2004]; between-groups anticipated shocks). The age dummies control for the age profile (within-groups anticipated shocks).[8] The specification proposed is the following:

$$lny_{it} = \beta_0 + \beta_1 YearDummy_{it} + \beta_2 AgeDummy_{it} + u_{it} \tag{3.2}$$

where lny_{it} is the natural logarithm of labour income.

The residuals u_{it} of the first stage regression represent an idiosyncratic indicator for labour income, cleansed from deterministic components. Under the unit root assumption, the first difference of this residual would represent the sum of a permanent and transitory shock [Browning et al., 2010]. However, allowing a certain degree of heterogeneity would potentially lead to a better specified measure of idiosyncratic risk. The paper assesses, using standard techniques, what underlying generating process best describes the *average* behaviour of individual incomes. The analyses show that individual labour income is likely to follow an ARMA(1,1) process with a unit root.[9][10] In the frame of these findings, the paper estimates a different AR(1) model on the incomes of each cross-sectional unit, allowing the parameters to be household specific.[11]

[7] Labour income is expressed in 2005 prices using an OECD CPI index. Labour income is considered after taxes. See Elmendorf and Kimball [2000] on the effect of labour income taxation on the demand for risky assets.

[8] Several cited contributions use 'potential experience' instead of age dummies (age, minus years of education, minus 5).

[9] The results of these analyses are contained in a dedicated appendix. Individual earnings (for the main and the second earner) follow an ARMA(1,1) process, with a unit root. The result is consistent with Abowd and Card [1989], Baker [1997], MaCurdy [1982]).

[10] It is debatable whether these techniques and this literature, mainly applied to individual earnings, can be applied to household earnings. Nevertheless, Jenkins [2000] mentions the earliest (and relatively most intuitive) methodologies as valid to study the covariance structure of household earnings (Abowd and Card [1989], Lillard and Willis [1978]). The paper decides not to follow this path, as in general the properties of household earnings, and the literature on household income dynamics, differ from those of individual earnings (see Jenkins [2000], Burgess and Propper [1998], Jalan and Ravallion [2001].

[11] It would be ideal not to restrict to AR(1) representations, but to estimate a different time-series process for each unit. This is simply not feasible where T is fixed (and small, as in this case). Likelihood based estimations of ARMA models need more observations to set initial values [Browning et al., 2010]. AR models of order greater than one would 'consume' too many degrees of freedom [Jenkins, 2000]. The measure proposed is still open to the critique of being miss-specified. This contribution aims at being a first step towards this direction.

The following equation describes the individual regressions performed:

$$u_{it_i} = \rho_{1i} + \rho_{2i} u_{it-1_i} + e_{it_i} \tag{3.3}$$

where the index t_i stresses the fact that the panel is not balanced. Both the intercept term and the AR(1) coefficient are individual specific, and allow for individual heterogeneity.[12] The labour income volatility measure would then be the standard deviation of the 'second stage residuals' of the estimated stochastic process, calculated for each individual.[13]

Recalling from equation (3) that e_{it} represents the residual from the individual specific regression, the labour income risk measure proposed in this paper is the following:

$$Risk = \sqrt{\frac{1}{t_i - 1} \sum_{t_i} e_{it_i}^2} \tag{3.4}$$

The methodology proposed does not fully solve the problems emphasized above. The lack of a long time span does not allow indeed to correctly specify each of the time series processes. This indicator, however, takes into account that the generating process could be stationary, specifies individual specific intercepts (see Baker [1997]), and allows each individual to have their *own* degree of idiosyncratic risk.

As anticipated, the focus of this paper is to analyze the influence of *household level* labour income risk. The procedure described in the present section is performed however on individual level income. More in detail, the labour income risk score is calculated for the main and the second earner, and for their joint risk (covariance). Calculating individual level risk allows to calculate household risk as well. Indeed, in the sample household income is approximately equal to the sum of main earners' incomes (see section 3.3). Calling y^{hh} household income, y^{ME} the income of the main earner, and y^{SE} the income of the second earner, it is clear that:

$$var(y^{HH}) \approx var(y^{ME}) + var(y^{SE}) + 2cov(y^{ME}, y^{SE}) \tag{3.5}$$

[12]A more detailed discussion on the distribution of the individual specific parameters is contained in Appendix 4.

[13]The residual represents indeed the difference between the actually observed value of income and its prediction, conditional on income in the previous period.

Household level risk is calculated using the above formula. This procedure allows also to study how the distribution of risk within the household influence their investment behaviour (see section 3.5.6.1). The following vector of risk scores is henceforth estimated:

$$\sigma = \begin{bmatrix} Risk_i^{HH} & Risk_i^{ME} & Risk_i^{SE} & Risk_i^{Joint} \end{bmatrix} \quad (3.6)$$

3.5 Estimations

The aim of the present section is to discuss the results of the econometric estimations performed. Section 3.5.1 delineates the estimation problem, framing it into economic theory and into the empirical literature on the topic. Section 3.5.2 describes the chosen estimation technique. Section 3.5.3 describes the model specification. Section 3.5.4 contains descriptive statistics of the regressors. Section 3.5.5 comments on the estimation results. Section 3.5.6 addresses the endogeneity issue and the hedging motive.

3.5.1 Theoretical background

The present contribution provides empirical evidence through reduced form estimation. Similar papers approach the problem in three different ways:

1. some papers study the mere participation choice in the risky assets market. They make use of binary choice models;

2. other papers treat the choice of not-investing as a legitimate investment choice (i.e. the household invests a 'zero amount'). Empirically, this represents a 'data censoring' problem (as defined in Miniaci and Weber [2002]). In this case, empirical analyses make use of single-index models such as Tobit;

3. finally, several studies divide the participation choice from the allocation one under. They assume that the sample of those who invest and the one of those who do not invest are drawn from different populations. Embracing this third alternative would lead to the estimation of sample selection models, like two-stage Heckit [Heckman, 1979].

Haliassos and Bertaut [1995], Bertaut [1998], King and Leape [1998] follow the first strategy. Guiso et al. [1996] and Chang et al. [2013] follow instead the second approach. Sample selection estimations are by now the most popular alternative. Perraudin and Sorensen [2000] and Betermier et al. [2012] implement a sample selection approach. Arrondel and Calvo-Pardo [2014] follow both the second and the third alternatives. Bonaparte et al. [2014] follow the three of them.

The present contribution focuses on the second one. The reason is twofold. Firstly, studying the mere participation choice does not inform on how the regressors influence the allocation choice. Secondly, although sample selection estimation would meet the purposes of the paper, it relies on restrictive assumptions that are not met in the data (above all, normality).[14]

3.5.2 The estimation technique

Before choosing what single-index technique to implement, one must carefully observe the distribution of the dependent variable. This is defined as the share of risky assets over financial wealth.[15] As standard in the literature, the distribution is characterized by a mass point at zero and a long tail. Most households do not invest at all in risky assets. The majority of those who invest dedicate a low percentage of their wealth to them. Few individuals invest instead a relevant fraction of their wealth. In addition, the dependent variable presents both a lower and an upper bound, as it is expressed in shares. The following graph shows its distribution:

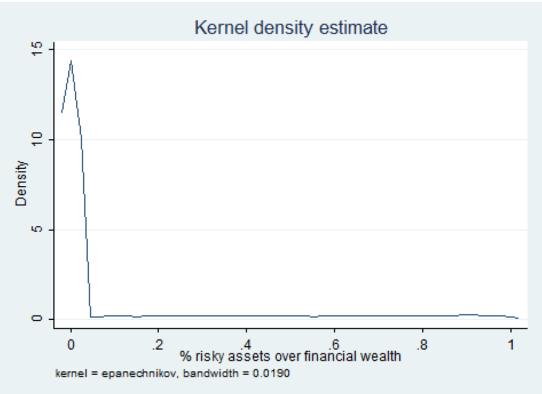

FIGURE 3.1: Distribution of the dependent variable

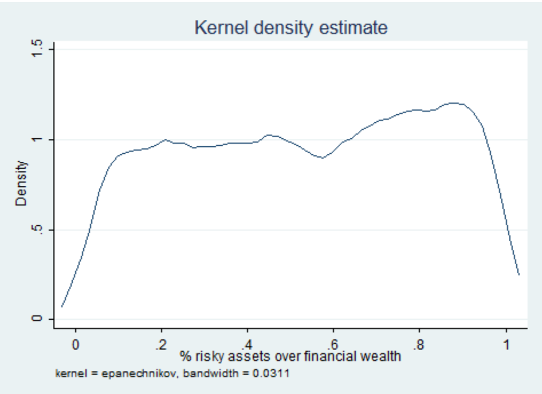

FIGURE 3.2: Distribution of strictly positive observations

[14]Estimation results for a participation choice model and for a two-stage Heckit model have been performed as a robustness check. The participation model results are displayed in appendix 2 (table 3.NUMBER), and are overall consistent with the main findings. The Heckman two stage estimation results are described in section 3.5.5.2.

[15]Risky assets are defined as: mutual funds, equity, shares in private limited companies, foreign securities, loans to cooperatives. Financial assets are defined as: risky assets, plus bonds in private companies, plus government bonds, plus liquidity.

Needless to say, OLS regression in this case would lead to completely unreliable results [Santos Silva et al., 2014]. As remarked above, several authors estimate models of this kind using Tobit [Guiso et al., 1996]. Tobit, however, is suitable for censored distribution problems. Censored distributions arise because of partial observability. A common example is when the dependent variable takes both positive and negative values, but zeros are observed instead of the negative ones. Tobit is not appropriate in this case because the zeros herein observed are *actual* zeros. The mass point does not arise because of partial observability. Individuals literally *choose* not to invest in risky assets. The data are not censored, but instead characterised by a corner solution.

Count models (like Poisson Pseudo ML) represent a more appropriate estimation alternative [Santos Silva and Tenreyro, 2006]. Albeit this choice is preferred to Tobit, Poisson Pseudo ML is not optimal if the dependent variable presents an upper bound (as in the present case). Santos Silva et al. [2014] suggest a valid alternative for data showing the characteristics herein presented.

In their contribution, Santos Silva et al. [2014] regress the number of sectors two countries trade in on a set of covariates. In analyzing this problem, they notice that:

1. the distribution of the dependent variable shows a mixed distribution with a mass point at zero (as couples of countries generally do not trade in most of the sectors);
2. the dependent variable shows an upper bound, as the number of sector is not infinite.

The authors propose an novel estimation technique that takes these features into account (the so-called 'flexible' model: Santos Silva et al. [2014]). More in detail, they propose the following modification of an exponential function:

$$E[A_{it} \mid x_{it}] = 1 - (1 - \omega e^{x_{it}\beta})^{\frac{-1}{\omega}} \tag{3.7}$$

where A_{it} is a variable bounded between zero and one, and ω is a shape parameter that adjusts the skewness of the distribution (left-skewed if $\omega < 1$, right-skewed if $\omega > 1$). Coefficients do not have a meaningful interpretations in this technique, but computation of marginal effects is possible. Santos Silva et al. [2014] show that their methodology fits the data better than previously applied methods (OLS, Tobit, Poisson Pseudo ML) and enhances the efficiency of the estimator.

Given that the characteristics of the dependent variable closely resemble the ones described in Santos Silva et al. [2014], the flexible model is chosen as an estimation technique. The paper also implements a novel test [Santos Silva et al., 2015] to show that the flexible model is actually preferred to a Heckman two stage estimation (see section 3.5.5.2).

3.5.3 The model specification

The present subsection describes the specification proposed and the regressors choice. The regression model proposed is the following:

$$\frac{A}{F}_{it} = g(\sigma_i\beta + x_{it}\gamma + x_{it}^{ME}\gamma^{ME} + x_{it}^{SE}\gamma^{SE} + z_i\delta + w_t\rho + \varepsilon_{it}) \quad (3.8)$$

where:

- $\frac{A}{F}_{it}$ is risky asset A divided by F, financial wealth[16];
- $g()$ is the functional form described in the previous section (the 'flexible' specification);
- vector σ_i contains the labour income risk measure(s). Section 3.4 details about the construction of this vector. Labour income risk is overall expected to be significant and to have a negative sign;
- vector x_{it}^{HH} contains household level variables, such as financial liabilities, real wealth, number of children;
- vectors x_{it}^{ME}, x_{it}^{SE} contain individual characteristics (age, sex, education) and labour market indicators attributable respectively to main and the second earner (occupation status, weekly worked hours). Moreover, they contain a labour income second-order polynomial, worked hours (per week), occupation status. Labour income is the natural logarithm of individual real labour income, deflated using an OECD CPI index;
- vector z_i contains geographical indicators (area of origin, size of the town of origin). This vector is intended to capture geographical or cultural differences in investment behaviour. Italy is characterized by deep north / south diversities. The northern part is more dynamic and overall richer, whilst the south is traditional and more agricultural. Italy also shows deep cultural differences between provinces and cities, captured by the town size dummies;

[16] See the previous subsection for a more detailed definition of the dependent variable.

- vector w_t contains year dummies and regional unemployment rate (disentangled by gender). Year dummies are intended to capture business cycle effects. The unemployment rate, taken from the national statistics institute (ISTAT) is a regional indicator disentangled by gender. This variable is expected to have a negative sign.

Similar specifications have been proposed, to cite the most recent contributions, in Betermier et al. [2012], Knupfer et al. [2013], Bonaparte et al. [2014].

3.5.4 Descriptive statistics

The aim of the present subsection is to provide descriptive statistics of the variables chosen for the estimation model. The following table displays the estimated values:

Table 3.1: Descriptive Statistics

Variable	Mean	Median	Std. Deviation
Female	1.09	1.00	0.29
Age (ME)	44.67	45.00	7.65
Education (ME)	3.83	4.00	0.78
Income (ME)	21944.50	19999.99	10787.76
Weekly worked hours (ME)	38.44	40.00	6.96
Boss (ME)	0.18	0.00	0.38
Blue Collar (ME)	0.29	0.00	0.46
In agriculture (ME)	0.02	0.00	0.13
In manufacturing (ME)	0.28	0.00	0.45
In finance (ME)	0.05	0.00	0.22
In public sector (ME)	0.48	0.00	0.50
Age (SE)	42.26	42.00	7.50
Education (SE)	3.92	4.00	0.74
Income (SE)	17060.01	16881.59	6936.35
Weekly worked hours (SE)	33.52	36.00	8.71
Boss (SE)	0.08	0.00	0.27
Blue Collar (SE)	0.26	0.00	0.44
In agriculture (SE)	0.02	0.00	0.15
In manufacturing (SE)	0.21	0.00	0.40
In finance (SE)	0.04	0.00	0.19
In public sector (SE)	0.55	1.00	0.50
Number of children	1.55	2.00	0.85
From the south	0.22	0.00	0.41
From the north	0.61	1.00	0.49
From a small town	0.10	0.00	0.30
From a big city	0.08	0.00	0.27
Own the house	0.78	1.00	0.42
Financial Liabilities	10155.97	0.00	24844.76
Financial Exposition	0.10	0.00	0.44
% risky assets over financial wealth	0.19	0.00	0.30
Observations	1971		

Note. *Age* is measured in years. *Education* is a discrete ordered variable, bounded between 1 and 6, where 1 is primary school, and 6 is postgraduate education. *Female* is a binary indicator, equal to one if the person is male, equal to two if female. *Worked hours* are measured in hours per week. *Labour Income* is real annual labour income in , deflated using a CPI index, considered after taxes. *Number of Children* is measured in units. *Boss, Bluecollar, South, North, Big Town, Small Town, Own the House* are binary indicators, equal to one if the individual is in the category identified, equal to zero otherwise. *Financial Wealth over Total wealth, Financial Exposition over Total Wealth, Share of Risky Assets over Financial Wealth* are shares, calculated as described in the variable name.

The main earner is on average older than than the second one (45 years against 42.17), and male in most of the cases (as female is equal to two, male is equal to one). On average, both the main and the second earner do not have a high school degree. The main earner generally works more hours per week (38.44 against 33.52) and usually has managerial positions (18% for the main earner, against 8% for the second earner). The households considered have on average 1.5 children, mainly live in the north (60% of the sample) in a medium sized town. Their wealth is mainly concentrated on housing (77%). Their level of financial exposition is about 10% with respect of their total wealth. 19% of their financial wealth is invested in risky assets.

3.5.5 Estimation results

The present section describes the regression results.

3.5.5.1 Main results

Table 3.2 contains the main estimation results. For clarity of exposition, table 3.2 shows the magnitude and the significance of the labour income risk measure only. Table 3.3 displays marginal effects of other regressors contained in the richest specification (table 3.2, column (8)). Several dummy variables coefficients are not displayed, given their conspicuous number, and are available upon request.

TABLE 3.2: Flexible Regression Model
Dependent Variable: Percentage of Risky Assets Over Financial Wealth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Risk (Household)	-0.199*	-0.749**	-0.093	-0.097	-0.085*	-0.080	-0.120**	-0.132**	-0.132**
	(-1.875)	(-2.530)	(-1.305)	(-1.353)	(-1.760)	(-1.618)	(-2.017)	(-2.322)	(-2.122)
Personal Characteristics	-	x	x	x	x	x	x	x	x
Geographical Indicators	-	-	x	x	x	x	x	x	x
Labour Market Indicators	-	-	-	x	x	x	x	x	x
Income	-	-	-	x	x	x	x	x	x
Wealth Indicators	-	-	-	-	x	x	x	x	x
Year Dummies	-	-	-	-	-	x	x	x	x
Unemployment	-	-	-	-	-	-	x	x	x
Sector of Activity	-	-	-	-	-	-	-	x	x
Omega	-6.962	-3.817	2.107	-0.677	-0.767	-0.767			
Observations	1,971	1,971	1,971	1,971	1,971	1,971	1,971	1,971	1,971
R- squared	0.010	0.041	0.080	0.087	0.093	0.094	0.183	0.195	0.195

Note. z-statistics in parentheses. Standard errors clustered by household. Marginal effects refer to a unit increase in the standard error of log-labour income. The last column contains specification (8), bootstrapped with 200 replications. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

More in detail, table 3.2 shows that a higher idiosyncratic human capital risk leads to the choice of investing less in risky financial assets. The significance is robust to an increasing number of regressors. Standard errors are clustered by household. A unit increase in the standard deviation of log-labour income is associated with a decrease in the dependent variable of about

13.2% (table 3.2, column (8)).[17] Column (9) contains the coefficients estimated in column (8), and standard errors taken from a bootstrapped variance covariance matrix.[18] The significance of the main coefficient is robust to the resampling procedure.

Table 3.3 displays marginal effects for a selection of remaining regressors.

All coefficients display the expected signs. The main earner's education positively influences risky assets allocation. A person with a postgraduate degree would invest 3% more than a person with a high school degree. This result is consistent with Guiso and Jappelli [2005], who find that better educated individual have a higher awareness of financial assets, and are as a consequence more likely to invest a share of their wealth in them. More children induce households to invest less in risky financial assets: 2 more kids are associated with a decrease in risky financial assets of 1.6%. Northern households invest more than southern ones: the gap between the two is 8%. Household living in small towns invest on average 2.4% less than other households. No significant correlation is instead found for living in a bigger city. The income of the main earner is positively correlated with investment in risky assets. The second earners worked hours are negatively correlated with the dependent variable: 10 hours more of work are associated with a decrease in allocation by 1%. Home-owners invest on average more (3.8%). The unemployment rate shows a negative and significant correlation (at the 1% level). A one-percent increase in the unemployment rate is associated with a 4% decrease in the dependent variable. This result is consistent with the hedging motive, as a higher unemployment rate indirectly represents a higher uninsurable background risk.

The influence of age on the demand for risky assets has been highly debated in the literature. Some scholars predict that younger individuals should invest more in risky assets [Gomes and Michaelides, 2005]. Other scholars conclude the opposite. King and Leape [1998] claim that individuals should invest in risky assets later in their life, as information acquisition is costly. In reduced form estimations, it is standard practice to control for non-linear effects of age (Guiso et al. [1996], Arrondel and Calvo-Pardo [2014], Bonaparte et al. [2014]). A specification including quadratic age, both for the main and the second earner, gives similar results to the ones displayed in tables 3.2 and 3.3. The risk measure has a significant, negative coefficient. The age variables do not show any significance. Both the second order age variables show a negative sign, suggesting a concave relationship between age and allocation in risky assets.

[17] All estimation tables report marginal effects. The coefficients of the risk measures would have had a more straightforward interpretation if the risk measures were expressed in levels. Unfortunately, when the volatility measures were expressed in levels, the convergence process was continuously prevented by the presence of discontinuous regions. Rescaling the risk measure and changing the optimization algorithm did not help the convergence of the likelihood function. The choice of using logarithms is driven by this fact.

[18] Estimation of the labour income volatility measure has been replicated 200 times together with the reduced form models. Resampling took place at a cross-sectional unit level. See Kapetanios [2008] for a discussion of such bootstrapping procedure when both T and N are large. Kapetanios [2008] also justifies the adoption of such procedure when N is large but T is fixed, as in this case.

TABLE 3.3: Flexible Regression Model Dependent Variable: Percentage of Risky Assets Over Financial Wealth

Specification of table 3.3, column (8)	
Age (ME)	0.000
	(0.687)
Female (ME)	0. Education (ME)
0.015***	
	(3.651)
Age (SE)	0.001***
	(3.014)
Education (SE)	0.007
	(1.587)
Number of children	-0.008***
	(-3.283)
South	-0.052***
	(-5.422)
North	0.028***
	(5.141)
Small Town	-0.024***
	(-3.608)
Bog Town	0.001
	(0.216)
Worked Hours (ME)	-0.000
	(-1.283)
Worked Hours (SE)	-0.001***
	(-2.886)
Income (ME)	-0.245***
	(-5.049)
Income (SE)	0.017
	(0.316)
Income squared (ME)	0.015***
	(5.403)
Income squared (SE)	-0.000
	(-0.048)
Own the house	0.038***
	(7.380)
Financial Liabilities	-0.000
	(-1.543)
Unemployment rate	-0.004***
	(-4.711)
Agriculture (ME)	-0.202
	(-1.214)
Manufacturing (ME)	-0.011
	(-0.370)
Banking (ME)	0.056
	(1.567)
Government (ME)	0.009
	(0.353)
Agriculture (SE)	0.087*
	(1.742)
Manufacturing (SE)	0.019
	(0.703)
Banking (SE)	0.001
	(0.015)
Government (SE)	-0.033
	(-1.429)

Note. z-statistics in parentheses. Standard errors clustered by household. Marginal effects refer to a unit increase in the variable level. *** p<0.01, ** p<0.05, * p<0.1

3.5.5.2 Robustness check: alternative techniques

Table 3.4 contains models estimated with alternative econometric techniques (OLS, Two-stage Heckit, Poisson Pseudo-ML, Tobit). The specification implemented is the one of table 3.2, column (8). Estimating the model with different techniques does not affect the sign and the significance of the coefficient of interest. Labour income risk negatively influences the choice of investing in risky assets, and the finding is robust to the implementation of different techniques.

TABLE 3.4: Alternative Techniques
Dependent Variable: Percentage of Risky Assets Over Financial Wealth

Technique:	OLS (1)	Tobit (2)	Poisson PML (3)	Two-stage Heckit (4)
Risk (Household)	-0.089**	-0.255**	-0.648**	-0.089***
	(-2.294)	(-2.220)	(-2.280)	(-2.844)
Personal Characteristics	x	x	x	x
Geographical Indicators	x	x	x	x
Labour Market Indicators	x	x	x	x
Income	x	x	x	x
Wealth Indicators	x	x	x	x
Year Dummies	x	x	x	x
Unemployment	x	x	x	x
Sector of Activity	x	x	x	x
Observations	1,971	1,971	1,971	2,027
R- squared	0.163		0.180	
Inverse Mills ratio	-	-	- 0.114	
				(0.759)

Note. t-statistics in parentheses. Standard errors clustered by household. Standard errors of two-stage Heckit calculated on Bootstrapped coefficients (100 replications). *** p<0.01, ** p<0.05, * p<0.1

OLS estimation results are displayed for the sake of completeness. It is however fairly obvious that OLS regression does not represent a suitable alternative, given the non-linearity of the model on one side [Santos Silva et al., 2014] and the sample selection bias on the other [Guiso et al., 1996]. The risk measure is negatively and significantly correlated. According to the OLS estimate, a unit increase in the standard deviation of the log of labour income would reduce the dependent variable by 9%.

Despite its strong distributional assumptions [Guiso et al., 1996], Tobit is a popular alternative in the literature on the topic. In Tobit, regression coefficients have a similar interpretation with respect to OLS estimation. They represent marginal changes on the subsample of uncensored observations only. Among those who decide to participate, a unit increase in the regressor reduces the dependent variable by 25.5%. This magnitude is higher than the one found for OLS and the flexible model. Those who decide to invest seem to be more sensitive to labour income volatility than the whole pool of individuals. This finding is consistent with the hedging motive. It must be kept in mind, however, that Tobit estimation in this particular case would lead to inconsistent estimates [Santos Silva and Tenreyro, 2006].

Santos Silva and Tenreyro [2006] suggest to estimate the gravity model of trade using Poisson PML. Their main goal is to estimate a heteroskedasticity robust measure of elasticity. Under the assumption that $E[y_i|x] = e^{x'\beta}$, the authors show that Poisson PML can be efficiently applied even to non-integer and non-poisson distributed data, without having to be concerned about the numerous zeros in the dependent variable (see also Gourieroux et al. [1984]). The assumption on the conditional mean seems too restrictive to choose Poisson PML as the best option, given the double bounded nature of the dependent variable. Poisson PML represents however a valid alternative to compare to as a robustness check.

The third column of table 3.4 displays the estimation result for a Poisson PML model. The result is consistent with the ones displayed so far, in terms of sign and significance. The coefficient can be interpreted as a semi-elasticity and leads to the conclusion that a unit increase in the dependent variable reduces investment in risky assets by 65%. This result is not at all comparable with the ones displayed so far. Santos Silva et al. [2014] show however that estimating a double bounded model using Poisson PML might lead to considerable bias in the coefficients. The technique does not seem to be a good candidate for the problem analyzed in this contribution.

As anticipated, Heckamn's two-stage estimation [Heckman, 1979] is currently the most popular alternative in the literature. In principle, implementing this technique would be useful to investigate the stockholing puzzle. As Betermier et al. [2012] stress, individuals firstly choose whether to participate, and secondly decide how much to allocate in risky assets. However, the distribution assumption of Heckman two-stage estimator are not met in this kind of data. The technique relies indeed on normality and homoskedasticity. The fourth column of table 3.4 displays estimation results for a two stage Heckit. Identification is generally facilitated by an exclusion restriction. Following Arrondel and Calvo-Pardo [2014], who in turn invoke King and Leape [1998], education is introduced only in the selection equation. Assuming that the participation choice is explained only by information costs, education would influence the participation choice, but not the allocation one.

The coefficient of the risk measure is again significantly different from zero, and displays a negative sign. It is worth noticing that the inverse Mills ratio is not statistically different from zero. As Bonaparte et al. [2014] emphasize, this would imply that the sample of participants is randomly drawn from the population. This finding is at odds with Betermier et al. [2012], Bonaparte et al. [2014] and Vissing-Jorgensen [2002], but similar to what found in Arrondel and Calvo-Pardo [2014].

A recently proposed test statistic allows to check whether the flexible model is preferable to a two-stage estimation procedure [Santos Silva et al., 2015]. In the context of corner solution problems, the test allows to discriminate between single-index models (like the flexible technique) and double-index models (like two stage Heckit), and to suggest which alternative is better. Table 3.5 below contains the test results. The statistics do not allow to reject the model implied by the flexible specification, but allow to reject the sample selection one. This finding does not necessarily imply that the zeros and the strictly positive values are generated by the same mechanism. However, this result, together with the non-significance of the inverse Mills ratio, hints that the zeros and the strictly positive values are not generated by different mechanisms, at least in the sample considered.

Table 3.5: HPC test results

Null hypothesis	p-value
The flexible model is not preferable to the sample selection model	0.000
The sample selection model is not preferable to the flexible model	0.986

3.5.5.3 Robustness check: alternative dependent variables

As already stated, a higher uninsurable labour income risk reduces investment in risky assets. This statement is supported by theoretical predictions (Gomes and Michaelides [2005], Cocco et al. [1998]), and herein empirically confirmed (tables 3.2 and 3.4). However, if households facing a higher labour income risk reduce their share of risky assets, they must increase the share of wealth invested in some safer options. Theoretical models generally assume that individuals can choose between two assets only, a risky one and a risk-free one (see the already mentioned Gomes and Michaelides [2005] and Cocco et al. [1998], or for example Sanroman [2015]). Government bonds and liquidity are both assumed to be risk-free, and are usually not distinguished from each other. However, as Gomes and Michaelides [2002] mention, long term government bonds are sometimes considered risky assets (see for example Jappelli et al. [2001]).[19] This

[19] On this ground, Gomes and Michaelides [2002] extend a baseline theoretical model with two assets by introducing an intermediate risk asset calibrated using long-term government bonds data.

subsection provides evidence on this issue, confirms the results found and supports the hedging motive.

Table 3.6 contains models estimated using the flexible technique, but with alternative dependent variables. The specification is the one of of table 3.2, column (8). All dependent variables are shares of a different financial asset over financial wealth.

TABLE 3.6: Flexible Regression Model
Alternative Dependent Variables

Dependent variable:	% Gov. Bonds (1)	% Liquidity (2)	% Shares (3)	% LT gov bonds (4)	% ST gov bonds (5)
Risk (Household)	-0.036 (-0.682)	0.155** (2.041)	-0.112** (-2.518)	-0.023 (-0.738)	-0.005 (-0.143)
Personal Characteristics	x	x	x	x	x
Geographical Indicators	x	x	x	x	x
Labour Market Indicators	x	x	x	x	x
Income	x	x	x	x	x
Wealth Indicators	x	x	x	x	x
Year Dummies	x	x	x	x	x
Unemployment	x	x	x	x	x
Sector of Activity	x	x	x	x	x
Observations	1,971	1,971	1,971	1,971	1,971

Note. z-statistics in parentheses. Standard errors clustered by household. Marginal effects refer to a unit increase in the standard error of log-labour income. *** p<0.01, ** p<0.05, * p<0.1

The table shows that Italian households are more likely to insure their labour income risk by investing in liquidity rather than in government bonds. This statement is supported by the regression results in the first two columns. Labour income volatility does not show a significant correlation with investment in government bonds. A higher human capital risk, however, pushes household liquidity up. When facing a higher background risk, households prefer to diversify their portfolio by holding liquidity.[20] A unit increase in the risk measure increases holding in liquidity by almost 15.5%.

The dependent variable in column (3) of table 3.6 is the fraction of shares over financial wealth. This is a subset of the dependent variable of the previous estimations. Not surprisingly, labour income risk significantly lowers allocation in this asset. The magnitude is lower than the one found in table 3.2, column (8): 11%. Columns (4) and (5) investigate whether households make difference between long term and short term government bonds in terms of hedging behaviour.

[20]It is worth noticing that, in the SHIW database, financial assets are the sum of risky assets, government bonds and liquidity. As a consequence, the three shares should sum up to one.

Government bonds do not seem to play a diversification role in the sample considered, not even when disentangled by maturity.

3.5.5.4 Robustness check: risk distribution within the household

The distribution of volatility among household members can influence how households hedge their background risk when investing in risky assets. Suppose that two households have the same levels of *individual* earnings risks (the variances). Suppose that the first is characterized by a null covariance (i.e. household incomes are uncorrelated with each other), whilst the second shows a positive covariance. *Coeteris paribus*, the second faces a higher background risk than the first. How would these two households differ when allocating wealth to risky assets?

Table 3.7 shows the regression results when household risk is split into individual and joint risks components (see section 3.4.2).[21]

Table 3.7: Flexible Regression Model
Dependent Variable: Percentage of Risky Assets Over Financial Wealth

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Risk (ME)	-0.135	-0.097	0.037	0.018	0.081	0.081	0.018	0.017
	(-1.051)	(-1.033)	(0.322)	(0.188)	(0.989)	(1.007)	(0.252)	(0.237)
Risk (SE)	-0.207	-0.074	-0.045	-0.049	-0.060	-0.058	-0.113	-0.134*
	(-1.417)	(-0.916)	(-0.403)	(-0.612)	(-1.004)	(-0.970)	(-1.273)	(-1.656)
Covariance	-1.228***	-1.026***	-0.567	-0.518*	-0.371*	-0.357*	-0.479**	-0.441**
	(-2.882)	(-2.769)	(-0.682)	(-1.786)	(-1.823)	(-1.844)	(-1.962)	(-1.977)
Personal Characteristics	-	x	x	x	x	x	x	x
Geographical Indicators	-	-	x	x	x	x	x	x
Labour Market Indicators	-	-	-	x	x	x	x	x
Income	-	-	-	x	x	x	x	x
Wealth Indicators	-	-	-	-	x	x	x	x
Year Dummies	-	-	-	-	-	x	x	x
Unemployment	-	-	-	-	-	-	x	x
Sector of Activity	-	-	-	-	-	-	-	x
Omega	-6.213	-1.486	1.784	0.520	7.593	8.458	-0.790	-0.753
Observations	2,151	1,996	1,996	1,971	1,971	1,971	1,971	1,971
R- squared	0.014	0.044	0.084	0.089	0.095	0.095	0.186	0.198

Note. z-statistics in parentheses. Standard errors clustered by household. Marginal effects refer to a unit increase in the standard error of log-labour income. *** p<0.01, ** p<0.05, * p<0.1

[21] Betermier et al. [2012] follow a similar empirical strategy. The authors restrict their sample to households in which it is possible to identify a main and a second earner. However, they focus their attention on household level risk, without comparing household variance with disentangled volatility. Moreover, they do not introduce a measure of covariance in their contribution. The present paper instead provides a comparison between household level risk taken as a whole and disentangled by household members, providing a different empirical evidence.

This specification takes into account the whole household income volatility, and allows to investigate how the distribution of risk within the household influences the dependent variable. The only risk indicator characterized by a significant and negative coefficient is the covariance. The result is robust to the introduction of an increasing number of regressors. A one unit increase in the joint risk is associated with a 44.1% decrease in the dependent variable (table 3.7, column (8)).

The covariance is the only component of household risk that induces them to invest less in risky assets. If household earnings are positively correlated, a negative income shock would negatively affect both incomes. This would make intra-household income compensations harder to perform. The result that a higher joint portion of risk reduces household investment in risky assets hints for a hedging behaviour in household investment.

3.5.5.5 Small sample size

Recall from section 3.3 that all analyses are performed on a sample of households observed *at least* 5 times. Although this is a sufficient number of observations to estimate AR(1) models (see section 3.4.2), concerns might arise that the estimates of interest are sensitive to small sample size issues. No household can indeed be observed for more then 9 waves. As a consequence, none of the models estimated following equation 3.3 has more than 8 degrees of freedom.

Before actually addressing the problem, it is worth discussing what are the consequences of estimating the key regressor on small sample. Estimating the risk measure on a small sample would lead to measurement error in the generated regressor. This, in turn, would bias the coefficient of the flexible model estimation towards zero. The magnitude of the bias would be bigger, the bigger the measurement error. All these consideration follow from standard econometrics results (see for example [Wooldridge, 2010, page 74]).

As a robustness check, table [NUMBER] presents estimation results where the labour income risk measure is calculated using the procedure described in section 3.4, but on smaller samples. The rationale of these estimations is that risk measures calculated on smaller samples should be characterized by a higher measurement error. Their coefficients, in turn, should be more biased towards zero than the one of displayed in table 3.2. More in detail, in column (1) the regression results are estimated on a sample that omits the first observation for each individual. In column (2) the regression results are estimated on a sample that omits the first wave (1989). In column (2) the regression results are estimated on a sample of individuals observed 5 to 8 times. The specification is the one of table 3.2, column (9).

Table 3.8: Flexible Regression Model
Robustness to amall sample

	First Observation Omitted (1)	First Wave Omitted (2)	Observed 5-8 times
Risk (Household)	-0.113**	-0.095**	-0.097*
	(-2.174)	(-1.969)	(-1.684)
Controls	x	x	x
Observations 1,689	1,560	1,485	

Note. z-statistics in parentheses. Standard errors clustered by household. Marginal effects refer to a unit increase in the standard error of log-labour income. *** p<0.01, ** p<0.05, * p<0.1

The regression results are in line with the predictions: in all three cases the magnitude of the coefficient is less negative, that is more biased towards zero. This result suggests that measurement error due to small sample size should be taken into account, and that the actual magnitude of the coefficient would be higher is the income processes were estimated on a longer time span. Nevertheless, estimating the volatility measure on small samples does not alter the main result: labour income risk significantly reduces allocation in risky assets. Further reductions of the sample confirm that the result is robust.

3.5.6 Endogeneity

The volatility measure introduced in the reduced form estimations is an endogenous regressor. Endogeneity needs to be addressed to claim that human capital risk has a direct influence on household portfolio choices, and that these choices are motivated by a hedging motive [Betermier et al., 2012]. The following sections will tackle endogeneity through several channels.

No attempt will be made to predict the direction of the omitted variable bias.[22] The following analyses do not solve the endogeneity problem. Yet, they show that the results commented so far are robust to the introduction of several omitted indicators and overall consistent with hedging (see also section 3.5.5.3).

[22] **?** (page 62) describes omitted variable bias in the linear case. He shows that the sign of the bias generated by an omitted variable z correlated with a regressor x can be easily identified *only by assuming* that z is uncorrelated with all other regressors. If the assumption is relaxed, the sign of the bias can be assessed only by calculating partial correlations. Clarke [2005] critically reviews the literature on omitted variable bias. He quotes Yatchew and Griliches [1985] as one of the few attempts of discussing the role of omitted variable bias in non-linear models. He concludes (still quoting Yatchew and Griliches [1985]) that trying to predict the direction of the bias in the non-linear case is a fruitless exercise.

3.5.6.1 Attitude towards risk

Attitude towards risk is an omitted variable in the model. A risk averse individual is likely to invest less in risky assets than a risk prone one. However, a more risk averse individual is also more likely to work in a position characterized by low income fluctuations. Both the main regressor and the dependent variable would be correlated with this omitted indicator. To control for attitude towards risk and reduce the omitted variable bias, additional regressions introduce on the right-hand-side past values of household's investment in risky assets (loosely following Betermier et al. [2012]). Table 3.8 displays estimation results that are consistent with this logic. The model in column (1) conditions on the amount invested in risky assets when the household is observed for the first time. Columns (2) to (4) introduce lags of the dependent variable among the regressors.

Table 3.9: Flexible Regression Model
Dependent Variable: Percentage of Risky Assets Over Financial Wealth

	(1)	(2)	(3)	(4)
Risk (Household)	-0.061**	-0.053	-0.057*	-0.046
	(-2.557)	(-1.579)	(-1.888)	(-1.550)
Initial % of risky assets	0.459***			
	(8.572)			
% risky assets (t-4)	-	0.220***	0.096***	0.037
		(8.423)	(3.335)	(1.297)
% risky assets (t-3)	-	-	0.241***	0.146***
			(7.696)	(5.107)
% risky assets (t-2)	-	-	-	0.280***
				(11.189)
Personal Characteristics	x	x	x	x
Geographical Indicators	x	x	x	x
Labour Market Indicators	x	x	x	x
Income	x	x	x	x
Wealth Indicators	x	x	x	x
Year Dummies	x	x	x	x
Unemployment	x	x	x	x
Sector of Activity	x	x	x	x
Observations	1,861	1,879	1,821	1,775

Note. z-statistics in parentheses. Standard errors clustered by household. Marginal effects refer to a unit increase in the regressor. *** p<0.01, ** p<0.05, * p<0.1

The past shares of risky assets show positive and highly significant coefficients. A discrete unit increase in the initial value of risky assets is associated with a 45.9% increase in the share of risky assets. This implies that, for example, an increase in the initial investment by 5 percentage points would raise the dependent variable by 2.25% (table 3.8, column (1)). Columns (2) to (4) have a similar interpretation. In column (4), for example, a 5% increase in the amount of risky assets invested two periods before increase the amount invested today by 1.42%

In column (1), the risk measure preserves its sign and significance (at the 10% level). A unit increase in the risk measure reduces the dependent variable by 6.1% (table 3.8, column (1)). The magnitude of the coefficient is reduced with respect to table 3.2. The influence of labour income volatility is robust to controlling for attitude towards risk. Although consistent with the overall findings, the results contained in columns (2)-(4) are less robust than the ones of column (1), probably because of the bias generated by reverse causality.

3.5.6.2 Bargaining power

The distribution of bargaining power within the household represent another potential source of endogeneity. Suppose that two households ('one' and 'two', for simplicity) show the same level of labour income volatility. Suppose also that the second earner of household 'one' has less bargaining power than the second earner of household 'two'. *Coeteris paribus*, household 'two' would make safer investment choices than the other one, under the assumption that women (who represent the vast majority of second earners) are more risk averse.[23]

A specification including the share of income of the SE over household income has been estimated. The introduction of this additional regressor does not particularly influence the main results. Little support is found towards an influence of intra-household bargaining on investment choices. The coefficient of the risk measure is characterized by the same significance, same sign, and comparable magnitude with respect to the one of table 3.2, column (8) (-0.13, p-value: 0.04).[24] The coefficient of the bargaining indicator, although not significant, is negative. The result is consistent with the empirical evidence that women implement less risky investment behaviours.

[23]Empirical estimations show that females are generally more risk averse. The evidence on this issue is controversial, though (Jianakoplos and Bernasek [1998], Schubert [1999], among others). I refer to the cited papers for a debate on the topic.

[24]No table is reported for this estimation. The results are reported in appendix 2, table 3.NUMBER

3.5.6.3 Covariance income shocks - risky assets returns

The correlation between household labour income and asset returns influences their portfolio choices (Bonaparte et al. [2014], Arrondel et al. [2010]). Suppose that an individual is employed in a certain sector of activity, and that they decide to invest a share of their savings on stocks in the same sector in which they work. If a negative shock hits that particular sector, both human capital and stock market returns will be negatively influenced. As a consequence, a good strategy to insure oneself against human capital risk is to invest in assets the returns of which are negatively correlated with one's labour income. However, familiarity reasons may induce investors to participate in a sector the returns of which are positively correlated with their earnings [Massa and Simonov, 2006].

Table 3.9 provides evidence on this issue. Columns (1) restricts the analysis to households showing a positive covariance income - returns. Column (2) restricts the analysis to households showing a negative covariance income - returns. The specification is the one of table 3.2, column (8).

Table 3.10: Flexible Regression Model
Dependent Variable: Percentage of Risky Assets Over Financial Wealth

	$Cov(Inc.Shocks, Ret.) > 0$ (1)	$Cov(Inc.Shocks, Ret.) < 0$ (2)
Risk (Household)	-0.076* (-1.780)	0.173 (0.681)
Personal Characteristics	x	x
Geographical Indicators	x	x
Labour Market Indicators	x	x
Income	x	x
Wealth Indicators	x	x
Year Dummies	x	-
Unemployment	x	x
Sector of Activity	x	-
Observations	1,620	351

Note. z-statistics in parentheses. Standard errors clustered by household. Marginal effects refer to a unit increase in the standard error of log-labour income. To facilitate convergence, the specification of column (2) does not contain year and sector of activity dummies. *** p<0.01, ** p<0.05, * p<0.1

Consistently with the hedging motive, columns (1) shows a negative and significant correlation between labour income risk and household allocation in risky assets. When earnings and returns are positively correlated labour income volatility has a stronger negative influence on risky assets allocation than in the sample fully considered. This finding is consistent with the hedging motive. The magnitude of the coefficient is lower than the one in table 3.2, column (8) (-7.6%).

When income shocks and returns are negatively correlated, a positive and significant coefficient of the labour income risk measure would represent a striking evidence towards the hedging motive. In column (2) the labour risk coefficient is indeed positive and with a sizable magnitude (17.3%). It does not show any significance, though. Additional computations show, however, that participation in the stock market is much higher when the covariance income shocks-returns is negative (almost triple: 74.49%, against 27.86% when the covariance income-returns is positive). These results overall support the hedging motive.

3.6 Conclusions

The aim of the present paper is to analyze whether labour income risk negatively influences household portfolio allocation in risky assets. Reduced form estimations show that this is the case, consistently with economic theory and with the empirical findings on the topic. The result is robust to the introduction of an increasing number of regressors. Inference based on a bootstrapped variance covariance matrix confirms the result. Estimating the model using different, alternative techniques does not affect the main findings. Moreover, estimating the model using alternative dependent variables strengthens the evidence that labour income volatility crowds out allocation in risky assets.

Addressing endogeneity supports the idea that labour income volatility has a direct influence on household investment, and that households implement a hedging behaviour. Additional empirical results suggest that attitude towards risk and the covariance between household earnings' shocks and assets returns matter in explaining the negative relationship between human capital risk and household investment in risky assets. These findings generally support the hedging motive.

The reduced form models implement an original measure of idiosyncratic labour income risk that performs consistently with economic theory and previous empirical evidence. Not only the volatility measure coefficient shows the predicted sign, but its significance is overall consistent across different model. Finally, on a methodological ground, the technique chosen seems appropriate to estimate the empirical model. Statistical tests allow to conclude that this technique is preferable with respect to an alternative double-index model (two-stage Heckit).

Appendix 2: additional results

TABLE 3.11: Logistic Regression. Dependent Variables: holding risky assets.

	(1)
Risk (HH)	-0.133**
	(-2.005)
Age (ME)	-0.001
	(-0.138)
Education (ME)	0.039
	(1.365)
Age (SE)	0.004
	(0.895)
Education (SE)	-0.008
	(-0.251)
Number of children	-0.020
	(-1.038)
From the south	-0.032
	(-0.439)
From the north	0.024
	(0.537)
From a small town	-0.057
	(-1.001)
From a big city	0.013
	(0.253)
Weekly worked hours (ME)	-0.001
	(-0.654)
Boss (ME)	0.002
	(0.038)
Blue Collar (ME)	-0.065
	(-1.493)
Weekly worked hours (SE)	-0.004***
	(-2.671)
Boss (SE)	-0.027
	(-0.456)
Blue Collar (SE)	-0.049
	(-1.147)
Labour Income (ME)	-0.309
	(-0.668)
Labour Income (SE)	-1.051***
	(-2.781)
Labour Income Sq. (ME)	0.018
	(0.751)
Labour Income Sq. (SE)	0.061***
	(2.834)
Own the house	-0.015
	(-0.387)
Financial Liabilities	-0.000**
	(-2.029)
Unemployment rate	-0.027***
	(-3.415)
In agriculture (SE)	0.065
	(0.647)
In manufacturing (SE)	-0.011
	(-0.241)
In finance (SE)	-0.033
	(-0.448)
In public sector (SE)	-0.067*
	(-1.725)
In agriculture (ME)	-0.329
	(-1.583)
In manufacturing (ME)	-0.016
	(-0.335)
In finance (ME)	0.088
	(1.391)
In public sector (ME)	-0.016
	(-0.352)
Observations	1971

Marginal effects; t statistics in parentheses
(d) for discrete change of dummy variable from 0 to 1
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Appendix 3: The covariance structure of earnings

The present appendix develops an analysis of the covariance structure of earnings using the SHIW database, using standard techniques. The computations are performed to obtain a better understanding of individual labour income generating processes. The findings displayed provide empirical justification for the creation of the idiosyncratic labour income risk measure described in section 4. In particular, the evidence found justifies the specification of the second stage regression.

The procedure used to determine the covariance structure of earnings is well established in the literature on the topic (see for example, Abowd and Card [1989]). The analyses rely on the assumption that all units in the sample are characterized by the same income generating process. The following sections describe the estimation procedure, and display empirical results for household heads (ME). The analysis follows exactly the same steps for for the second earner's labour income. The results obtained are quantitatively and qualitatively comparable to the ones herein shown. As a consequence, evidence pertaining to the second earner is not displayed (although available upon request).

First stage regression

The first step of this analysis involves the estimation of a first stage regression of log - earnings on a set of covariates. What regressors to introduce into the first stage is still open to question. Several scholars (see for example Baker [1997]; Abowd and Card [1989]) regress the logarithm of labour income on potential experience, usually defined as individual age minus years of education minus 5. Year effects are generally introduced as well. Baker [1997] for example regresses the logarithm of labour income on potential experience, using a system of simultaneous equations in which each equation corresponds to an year. Browning et al. [2010] regress log of labour income on "on age and experience variables and time dummies".

The specification proposed in the paper is the following:

$$lny_{it} = \beta_0 + \beta_1 YearDummy_{it} + \beta_2 AgeDummy_{it} + u_{it} \tag{3.9}$$

which is the same described in section 4. The reasons for this choice are the same described in section 4, to which I refer to avoid redundancies.

The creation of a variance-covariance matrix

The residuals of the first-stage regression represent income processes cleansed from within and between group anticipated shocks [Abowd and Card, 1989]. To study the covariance structure of earnings, the literature suggests the creation of a variance-covariance matrix of the residuals. The estimation of this matrix is performed by pooling together individuals across waves. Calling $\bar{u}_i$ the between-group mean of the residuals, the deviations from the mean are defined by $\tilde{u}_i = u_{it} - \bar{u}_i$. The estimated variance-covariance matrix is consequently: $CL = \frac{1}{n_t} \sum_{i=1}^{n_t} \tilde{u}_i \tilde{u}_i{}'$, where CL stands for "covariance of income levels". CD (covariance of income deviations) is the name of the variance covariance matrix for Δu_{it}, estimated using the same formula.[25] The variance covariance matrices are informative about the nature of the income generating process. The following sub-subsections describe the empirical results for the process in level and in first differences respectively.

The process in levels

The following table displays the variance covariance matrix for the level of earnings (CL) of the main earner.

TABLE 3.12: Correlations - ME labour income (levels)

	2010	2008	2006	2004	2002	2000	1998	1995	1993	1991
2010	**0.198***	0.785*	0.680*	0.749*	0.676*	0.621*	0.359*	0.422*	0.035*	0.018*
2008	0.150*	**0.185***	0.731*	0.707*	0.668*	0.687*	0.412*	0.384*	0.422*	0.031*
2006	0.156*	0.162*	**0.266***	0.697*	0.660*	0.584*	0.401*	0.336*	0.393*	0.216*
2004	0.133*	0.121*	0.143*	**0.159***	0.743*	0.655*	0.455*	0.427*	0.418*	0.364*
2002	0.121*	0.115*	0.137*	0.119*	**0.161***	0.680*	0.541*	0.458*	0.436*	0.375*
2000	0.108*	0.116*	0.118*	0.102*	0.107*	**0.153***	0.700*	0.535*	0.527*	0.368*
1998	0.059*	0.065*	0.076*	0.067*	0.080*	0.101*	**0.136***	0.581*	0.499*	0.461*
1995	0.057*	0.050*	0.053*	0.052*	0.056*	0.064*	0.065*	**0.092***	0.584*	0.395*
1993	0.040*	0.070*	0.078*	0.064*	0.068*	0.079*	0.071*	0.068*	**0.148***	0.486*
1991	0.003*	0.038*	0.038*	0.050*	0.052*	0.049*	0.058*	0.041*	0.064*	**0.118***

Autocovariances below diagonal, autocorrelations above diagonal, cross sectional volatility on diagonal. * : significant at the 5% level.

All the correlations displayed are significant at the 5% level. The generating process is consequently characterized by a high persistence. This finding can be attributable to different reasons. The process can be characterized by non stationarity. Alternatively, the generating process can be stationary, and simply show a high persistence that does not die out in the time span considered. Moreover, if the process is stationary and invertible, and if the error term follows an MA(q) process, the Wold representation theorem predicts that the autoregressive part can be expressed as an AR(∞) model. This would lead to a highly persistent correlogram as

[25]The notation follows Abowd & Card's statistical appendix (1989), to which I refer for further details.

well. More detailed analysis of the process is consequently needed to draw accurate conclusions about the nature of the generating process. In particular, the very same correlation analysis should be performed on the process in first differences to understand what the nature of the error term is (Baker [1997];Meghir and Pistaferri [2004]).

The following table shows how cross sectional volatility evolves along time. The table displays how the sample size evolves, as the longitudinal data used are not balanced. Indeed, as the sample contains individuals observed more than five times, at the beginning of the time span the number of individuals has to be lower than in the middle of the period considered. Moreover, displaying how volatility evolves is informative about the stationary properties of the process. In a covariance stationary process, indeed, the variance should be constant along time. However, if the variance increases or decreases with time, this represent a hint for the presence of non stationarity.

Table 3.13: Cross sectional volatility across time - ME

Year	Variance	Frequency
1991	0.118	138
1993	0.148	170
1995	0.092	177
1998	0.136	228
2000	0.153	255
2002	0.161	256
2004	0.159	224
2006	0.266	210
2008	0.185	183
2010	0.198	140

The following graph plots cross sectional variation against time:

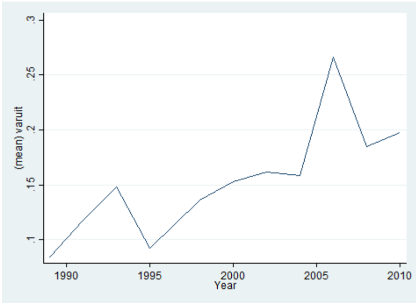

FIGURE 3.3: Cross sectional variation against time

A covariance stationary process is characterized by a constant variance along time. The graph shows instead a growing variance, or, in other words, a growing inequality. However, despite at an aggregate level the process seems to show non stationary properties, it is still early to conclude that the income generating process is non stationary. This pattern can be indeed due to individual heterogeneity.

In a covariance stationary process, the covariance merely depends on the lag order considered, and not on time. Hence the first order autocorrelation of a stationary process should oscillate around a certain mean if plotted against time. The following graph shows the described plot. The observation for 2010 represents the correlation between 2010 and 2008. The observation for 2008 represents the correlation between 2008 and 2006, and so on.

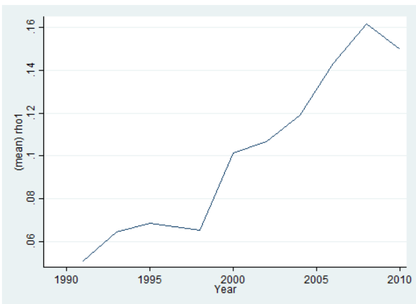

FIGURE 3.4: First order autocovariance plotted against time

As for cross sectional dispersion, first order autocorrelation shows a growing pattern. That is, the covariance does not seem to depend on the lag itself, but also on the time period considered. According to this graph, the process has non stationary properties.

The following two graphs plot second and third order autocorrelations against time. The conclusions drawn accordingly are similar to the ones drawn for the variance and the first order autocorrelation.

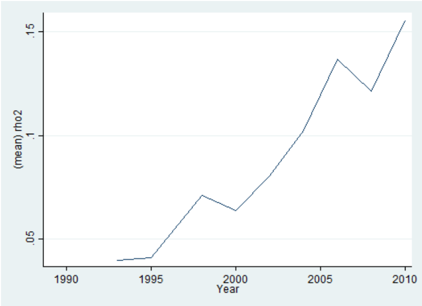

FIGURE 3.5: Second order autocovariances

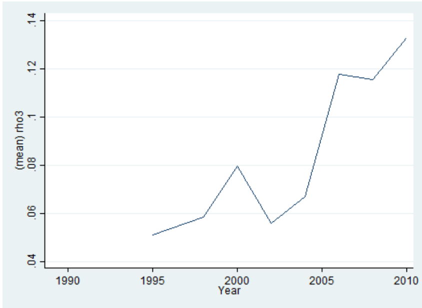

FIGURE 3.6: Third order autocovariances

In line with the time series plots of the variance and the autocovariances, the process shows not stationary properties. The variance depends on time, and the autocovariances as well. The following table contains OLS regression results where the dependent variables are the second moments, and year represents a regressor. In the following, and using the notation commonly used in time series analysis, the autocorrelation for the j^{th} lag is indicated as γ_j.

TABLE 3.14: Second moments regressed on time - ME

VARIABLES	(1) γ_0	(2) γ_1	(3) γ_2	(4) γ_3	(5) γ_4	γ_5
Year	0.00574***	0.00600***	0.00685***	0.00556***	0.00566**	0.00324
	(4.200)	(11.42)	(10.21)	(7.328)	(3.781)	(1.233)
Constant	-11.32***	-11.89***	-13.62***	-11.05***	-11.27**	-6.424
	(-4.150)	(-11.33)	(-10.15)	(-7.274)	(-3.754)	(-1.221)
Observations	11	10	9	8	7	6
R-squared	0.639	0.933	0.925	0.771	0.703	0.347

Robust t-statistics in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The regression results provides evidence that the second moments of the process are time dependent. This finding is consistent with the literature on the topic. Baker [1997] for example points out that "there is evidence of non stationarity as variances rise briefly at the beginning of the panel, and then trend rather randomly before rising sharply in the final years". Abowd and Card [1989] notice that "variances ... vary over time. Cross sectional dispersion in earnings was relatively small in 1972-73 and relatively large in 1975-1976". Variance dependence on time as a hint for non stationarity is an established result in the literature. The analysis displayed so far suggest that the generating process for earnings is non stationary.

The process in differences

The present subsection discusses the analysis of the first difference of the earnings process. This represents a well established procedure in the literature, as studying the correlation of a process in differences is informative about the time series properties of the error term (Baker [1997], Meghir and Pistaferri [2004]). Supposing indeed that the generating process has an AR(1) representation with a unit root and the error term shows an MA(q) representation, as indicated in the following equation:

$$y_t = c + y_{t-1} + \Psi(\varepsilon; q)$$

the correlogram of Δy_t is informative about the time series properties of the error term. Indeed, if the correlogram dies, say, after one lag, the error term has an MA(1) representation. If the correlogram dies abruptly after 3 lags, the error term has an MA(3) representation. Quoting Baker [1997]: "assuming (the presence of a unit root) ... all the autocovariance above the (MA) order should be equal to zero; that is , in contrast to (the autocovariances of the process in levels) there is no persistent serial correlation in the earnings growth rates." The following table displays the variance covariance matrix for the process expressed in first differences (CD).

TABLE 3.15: Correlations - ME labour income (first differences)

	2010	2008	2006	2004	2002	2000	1998	1995	1993	1991
2010	**0.079**	-0.239*	-0.114	-0.066	-0.077	-0.098	-0.155	0.099	0.035	1.521
2008	-0.026*	**0.156**	-0.696*	-0.008	-0.061	-0.024	-0.033	0.153	-0.080	-0.011
2006	-0.012	-0.103*	**0.141**	-0.275*	0.117	0.096	-0.035	-0.093	0.155	0.193
2004	-0.005	-0.001	-0.028*	**0.074**	-0.449*	-0.112	-0.034	0.168	-0.140	-0.048
2002	-0.005	-0.006	0.011	-0.030*	**0.062**	-0.380*	-0.050	0.074	-0.012	-0.057
2000	-0.007	-0.002	0.009	-0.008	-0.024*	**0.065**	-0.161*	-0.089	0.175	0.051
1998	-0.011	-0.003	-0.003	-0.002	-0.003	-0.011*	**0.067**	-0.284*	-0.093	-0.033
1995	0.008	0.017	-0.010	0.013	0.005	-0.006	-0.020*	**0.076**	-0.575*	-0.156
1993	0.026	-0.012	0.022	-0.014	-0.001	0.017	-0.009	-0.059*	**0.140**	0.784*
1991	0.094	-0.020	0.016	-0.003	-0.003	0.003	-0.002	-0.009	0.064*	**0.048**

Autocovariances below diagonal, autocorrelations above diagonal, cross sectional volatility on diagonal. * = significant at the 5% level.

The table shows how the correlogram dies after one lag, suggesting that the error term has an MA(1) representation. An MA(1) process in the error term represents a plausible explanation for the high persistency of the process in level, in case of stationarity and invertibility.

The following graphs plot cross sectional variation, first and higher order autocorrelations against time:

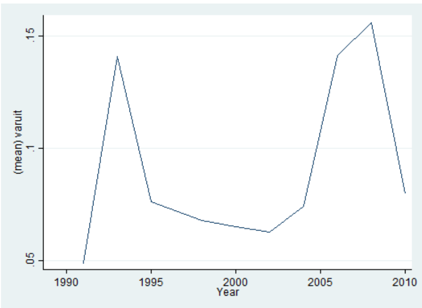

FIGURE 3.7: Cross sectional variation

FIGURE 3.8: First order autocovariance

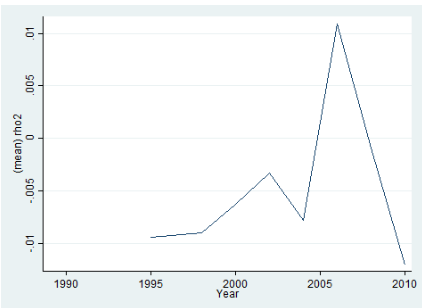

FIGURE 3.9: Second order autocovariances

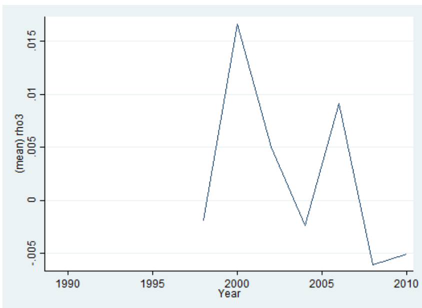

FIGURE 3.10: Third order autocovariances

The process expressed in first differences shows stationary properties, despite cross sectional volatility seems to be growing during in the last years of the sample. The following regression shows how the second moments do not actually depend on time.

TABLE 3.16: Second moments regressed on time - ME

	(1)	(2)	(3)	(4)
VARIABLES	γ_0	γ_1	γ_2	γ_3
Year	0.00197	-0.00207	0.000404	-0.000908
	(0.873)	(-0.962)	(0.731)	(-1.117)
Constant	-3.848	4.110	-0.814	1.821
	(-0.853)	(0.955)	(-0.737)	(1.117)
Observations	10	9	8	7
R-squared	0.106	0.153	0.081	0.218

Robust t-statistics in parentheses
*** p<0.01, ** p<0.05, * p<0.1

To conclude, the process in differences shows stationary properties, with an error term characterized by an MA(1) representation.

Conclusion

The analyses performed, under the (limiting) assumption that all individuals in the sample are characterized by the same process, leads to the conclusion that the income generating process in a non-stationary ARMA(1,1) process. This conclusion is consistent with the literature on the topic. The assumption that all individuals in the sample are characterized by the same process is partially relaxed in the paper, as labour income processes are estimated at an individual level (allowing them to have their own intercept and slope parameters). The estimation of AR(1) models, however, is empirically justified by the theoretical findings herein displayed. The degree of individual heterogeneity, could not be pushed further due to data unavailability.

Appendix 4: discussion on the estimation of equation 3.3

The present appendix contains the estimation results for equation 3.3, and aims to support the choice of estimating a individual specific models, rather than simply assuming that each individual in the sample has a unit root in their earnings (see Browning et al. [2010]). Equation 3.3 is reported below to facilitate the comprehension.

$$u_{it_i} = \rho_{1i} + \rho_{2i} u_{it-1_i} + e_{it_i}$$

Recall from section 3.4 that the above equation is estimated for the income processes of the main earner (ME) and the second earner (SE).

The following tables describe the distributions of the two parameters. Table 3.15 focuses on the main earner, table 3.16 focuses instead on the second earner. Kernel density estimators of the intercept and slope parameters provide graphical representations of the distributions.

TABLE 3.17: Distribution of regression parameters - ME

	(1) Sample Size	(2) Mean	(3) Std. Dev.	(4) α_{25}	(5) α_{50}	(6) α_{75}	(7) α_{90}
Intercept term (ρ_{1i})	919	0.18	4.53	0.09	0.26	0.48	0.75
Slope (ρ_{2i})	919	-0.03	4.29	-0.26	0.05	0.44	0.76
Slope (ρ_{2i}) between -1 and 1	836	0.07	0.42	-0.24	0.04	0.39	0.64

TABLE 3.18: Distribution of the intercept term - SE

	(1) Sample Size	(2) Mean	(3) Std. Dev.	(4) α_{25}	(5) α_{50}	(6) α_{75}	(7) α_{90}
Intercept term (ρ_{1i})	607	0.01	1.79	0.03	0.29	0.56	0.81
Slope (ρ_{2i})	607	0.07	2.68	-0.29	0.01	0.41	0.91
Slope (ρ_{2i}) between -1 and 1	526	0.02	0.41	-0.27	-0.01	0.29	0.61

With respect to the introduction of intercept terms, the canonical, unit root representation of earnings does not generally contain an intercept parameter (see appendix A in Browning et al. [2010]). This is true for both structural specifications modeling life-cycle investment decisions (such as Gomes and Michaelides [2002], Alan [2006]), and structural models of earnings dynamics (such as Meghir and Pistaferri [2004]). When an intercept term is included, it is generally intended to capture individual heterogeneity (see Sanroman [2015]). Empirical contributions evaluating the covariance structure of earnings usually remain agnostic on the introduction of an intercept term, as they generally estimate earning processes in growth rates (see Baker [1997], Abowd and Card [1989]). Does estimation of equation 3.3 support the introduction of an intercept parameter?

As the above tables display, the distribution of individual specific intercepts present averagely positive values (0.18 and 0.01 respectively), and high dispersion (consistently with Browning et al. [2010]). Moreover, the kernel density estimators shows that the low magnitude of these estimates (and in particular, of the estimator of the second earner) is likely to be driven by few, negative outliers. The distribution of the intercept terms is clearly skewed towards positive values. Empirical evidence points here towards the introduction of an intercept parameter. Following Browning et al. [2010], this can be interpreted as a 'long-run mean', that individual expectations take into account (see[Browning et al., 2010, Eq. 7]).

The key question, at this stage, is whether earning processes have a unit root or not. Consistently with Browning et al. [2010], the paper supports the idea that earning processes are highly heterogenous, and on average stationary (see tables 3.15 and 3.16). The magnitudes of the

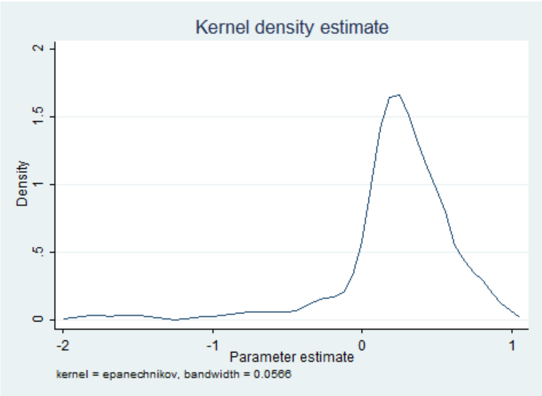

FIGURE 3.11: Distribution of the intercept terms - ME

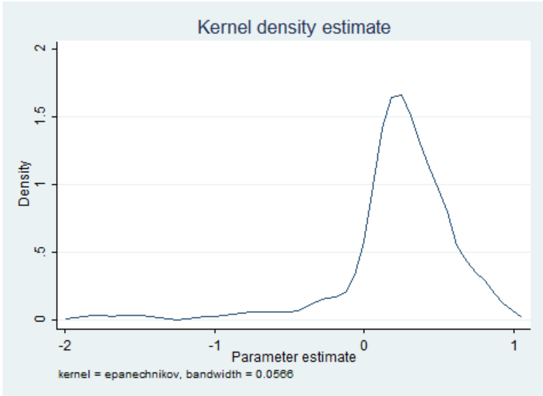

FIGURE 3.12: Distribution of the intercept terms - SE

slope parameters are however significantly lower than the ones found in Browning et al. [2010]. The authors, however, estimate a more flexible, and overall richer specification. Comparisons over the magnitudes would be hazardous at this stage. The kernel density estimators reported below provide a graphical representation of this findings.

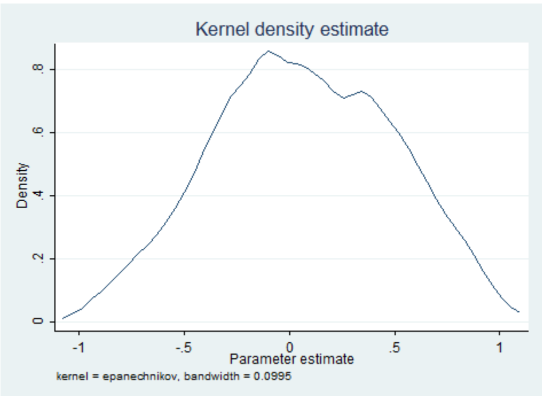

FIGURE 3.13: Distribution of the slope - ME

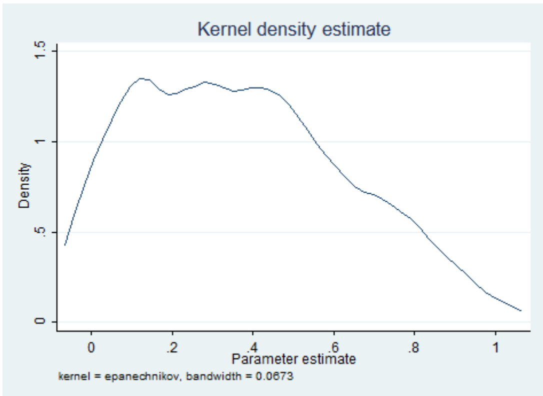

FIGURE 3.14: Distribution of of the slope for positive values- ME

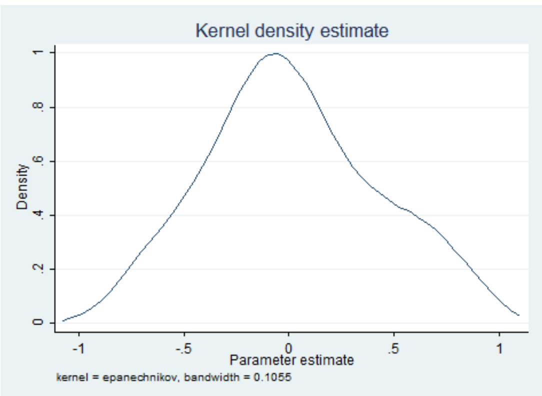

FIGURE 3.15: Distribution of the slope - SE

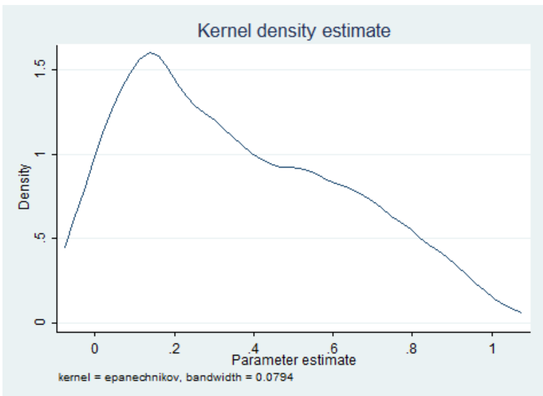

FIGURE 3.16: Distribution of of the slope for positive values- SE

To conclude, this appendix discusses whether the estimation of the model described by equation

3.3 supports the canonical unit root representation [Meghir and Pistaferri, 2004], or the emerging idea that income processes should be characterized by more heterogeneity [Browning et al., 2010]. The empirical evidence provided supports the latter view: individuals have on average a stationary earnings process, with a positive intercept. This result is at odds with the canonical earnings representation.

Chapter 4

Interest Rates and Default in Subprime Credit Markets: Evidence from a Randomized Price Experiment (With Sule Alan and Eric Smith).

Summary

Exploiting an exogenous variation generated by a subprime lender's randomized price experiment, we estimate the causal effect of cost of borrowing on default behavior. After segmenting the experiment sample into different risk categories based on its internal scoring algorithm, the lender randomly allocates the individuals within each segment into treatment and control groups. A six-month follow-up period allows us to estimate the effect of a 5 percentage point increase in interest rate on default behavior conditional on risk characteristics of the borrowers. We find that a large interest rate increase significantly raises the probability of default for liquidity constrained individuals. We find that the effect sizes are heterogenous across risk groups. Robustness checks confirm the main results.

4.1 Introduction

Access to high cost credit and its potential impact on household financial well being has been a topic of contentious debates. Fueled by the latest financial crisis, subprime credit markets are central to these discussions. Supporters argue that subprime credit markets give access to credit to otherwise rationed costumers [Adams et al., 2009]. Although expensive, high cost credit provides insurance to temporary disruptions in households income (such as temporary lay offs or sickness leaves), and unanticipated expenditure shocks (such as car repairs and funerals). OFT [2010] for example reports that unemployed individuals are a relevant fraction of the sub-prime borrowers. High cost credit is also found to stimulate regional growth in developing countries where credit rationing is severe and the informal sector is prevalent (see Karlan and Zinman [2009] for a description of high-cost credit in South Africa).

On the other hand, skeptics of the high cost credit argue that it pushes poor and financially unsophisticated consumers disproportionately into debt trap and eventually tips them into delinquency or default. OFT [2010] documents that UK high-cost credit clients are on average poorer, financially illiterate and lack credit history.[1] PFRC - Bristol [2013] completes the picture by showing that vulnerable categories, such as young, disabled and uneducated individuals, or lonely parents, are over-represented among high cost borrowers. BIS [2013] shows that costumers in this segment are characterized by disproportionately large levels of default with respect to the regular market (above 20%). Adams et al. [2009] document that costumers is the U.S. subprime credit market for used cars default in over 50% of the cases. Similar evidence in found in Meier and Sprenger [2009]. From this point of view, high cost credit can generate significant social losses.

Household financial well-being is not the only concern related to high default rates in sub-prime credit markets. Defaulters are also potentially harmful for lending institutions. Theoretically, a profit-maximizing lender faces a trade-off when choosing what interest rates to apply to their clients [Stiglitz and Weiss, 1981]. High interest rates can generate high revenues. Yet, they might lower the borrower's incentive or capability to repay their debt, and eventually cause them to default. Gross and Souleles [2002] mention indeed that defaults increase the marginal cost of lending. Following these premises, it is clear that investigating the relationship between cost of credit and individual default is of great importance both for the demand and the supply sides of the market [Meier and Sprenger, 2009].

Albeit clear theoretically, the causal relationship between interest rates and default behaviour is empirically hard to address because of endogeneity issues. Indeed, observed variations in the cost of credit are generally endogenous to unobserved borrower characteristics. Vissing-Jorgensen [2010], for example, regresses loss rates (amount of not repaid debt over total debt) on a set of regressors *and* interest rate. She mentions that the interest rate reflects the lender's information on clients' expected future losses, implicitly stating that the regressor is endogenous

[1]Financially illiterate costumers make poorer borrowing choices and pay on average more fees [Lusardi and Tufano, 2009]. As a consequence, we argue, they are likely to default more.

to information not observed by the researcher. In this paper we make use of a unique database to estimate the causal relationship between cost of credit and default behaviour. Data consist of detailed credit card transactions from a private lender that serves exclusively the subprime market in the United Kingdom.[2] Besides all individual transactions and credit terms, the data contain information on a *randomized* price experiment implemented by the lender in February 2008.

The experimentation was performed on the *existing* clients as part of the lender's risk pricing practice so that it is *not* a solicitation based market experiment. The lender carried out the experiment using a block design where a randomly selected sample of existing customers was assigned to categories ('cells' from now onwards). The assignment was based on the customers' utilization rates (statement balance over credit limit) and internally developed credit scores. After allocating the experiment sample into different cells in November 2007, the lender randomly selected a number of individuals in each cell and allocated them into control groups. The treatment groups (remaining individuals) received a 5% interest rate increase that applies to their existing stock of credit card debt. By design, the interest rate increase under the treatment is exogenous to individual characteristics. Clients are followed for 6 months after the treatment (from February 2008 to July 2008), allowing us to observe default and identify the causal effect of the interest rates increase on default probability and credit demand.[3]

The block design used by the lender also allows us to assess whether the effect of the interest rate on default propensities and credit demand is *heterogeneous* across borrower types *within* the subprime population.[4] The results suggest that there is a significant heterogeneity in propensity to default and demand sensitivity among subprime credit users with respect to the interest rates increase. Low risk borrowers exhibit a negative treatment effect on the demand for credit: treated individuals reduce their credit demand 57% more than in comparison to individuals in the control in reaction to the interest rate increase. Consistently, these costumers do not show a treatment effect on default probability. Medium risk borrowers display an opposite behaviour. We are able to isolate a positive, sizable and significant treatment effect on default probability for this category (45% more than the control group, with a p-value of 0.013). As expected, these costumers do not show a significant treatment effect on credit demand. Finally, high risk borrowers show a very similar behaviour with respect to medium risk borrowers. They are characterized by a positive treatment effect on default probability. The treatment does not show significance, but we are able to show that this happens because of power issues.

[2]For confidentiality reasons, we do not disclose the name of the company. We will refer to it as the 'lender' from here on.

[3]The reason why the lender conducts this experiment and many other similar ones is to find out price sensitivities as part of their risk-based pricing practices

[4]This approach has been followed in other studies. Gine et al. [2012], for example, assess whether dynamic incentives have heterogeneous effects across predicted repayment groups.

We also contribute to the literature by assessing whether the treatment influences the lender's returns, and whether the experiment is an overall profitable intervention. Although the treatment causes low risk individuals to reduce their credit demand and high and medium risk individuals to default, we find a positive and significant effect on interest rates charges on almost all cells. Low risk individuals reduce their credit demands, however the price increase compensates the quantity reduction. Similarly, for higher risk individuals, the higher revenues generated by the price increase compensate the higher costs generated by individual defaults.

Several contributions provide evidence on the relationship between the cost of credit and households financial well being, both in developed and developing countries. The literature, and in particular the one implementing field experiments, is currently skewed towards the latter ones. Starting from advanced economies, Adams et al. [2009] show that in the U.S. sub-prime credit market for used cars, a one percent increase in the annual interest rate positively influences default rate by 2.2%. Ausubel [1999] finds that in the U.S. credit card market, inferior terms (higher pre-approved interest rate, shorter term) have a positive, significant influence on individual delinquency and charge-offs. Given the peculiar nature of the data he analyses (randomize trials), the author is able to attribute these findings to the presence of adverse selection. Agarwal et al. [2010] show that individuals who accept lower credit conditions are more likely to default, and emphasize the negative role of adverse selection on financial well-being. A study related to ours shows that low risk borrowers who fully utilize their credit limit lower their credit due to an exogenous increase of the interest rate [Alan and Loranth, 2013].

Contributions in development economics often involve well-crafted field experimental designs. Karlan and Zinman [2010] display positive evidence that high cost credit enhances financial well-being. The authors show that even at rates as high as 200%APR, access to credit produces net benefits for (marginal) borrowers over a range of outcomes in South Africa. They also find some evidence that the marginal loans are profitable. In another contribution, Karlan and Zinman [2009] identify informational asymmetries in the cash loan market in South Africa, showing that 13% to 21% of defaults is attributable to moral hazard. Information asymmetries have an important role in explaining the low financial well-being of sub-prime borrowers. Gine et al. [2012] make use of a randomized field experiment to show that correctly identifying borrowers' leads to higher repayment rate for high-risk clients in an unsophisticated credit market (rural credit in Malawi).

A related literature analyses the recent increase in personal bankruptcies in the United States. Livshits et al. [2010] analyze the possible causes for the observed increase in personal bankruptcies observed in the U.S. in the years between 1970 and 2010. Gross and Souleles [2002] estimate duration models of default probability, and attribute the recent and sizable increase in personal filings in the U.S. to the decrease in filing costs. White [2007] describes this recent U.S. trend in great detail, provides several explanations for it, and extensively reviews the literature on the topic.

The remainder of the paper is organized as follows. Section 4.2 motivates the outcome variables. Data and the experimental design are detailed in section 4.3. The empirical results are outlined in section 4.4. Section 4.5 concludes.

4.2 Theoretical motivations

The present section develops a simple theoretical framework to explain how the experimental design can influence the probability of default, using a two-period model. This simple, two-period specification will then be used to motivate the outcome variables of the empirical analysis (see 4.2.2). The model draws conclusions on the optimal individual lending behaviour, taking into account the characteristics of the experimental design.

The specification proposed is an intertemporal utility optimization problem, with two periods and risk averse individuals (logarithmic utility). The two period specification with risk averse agents is consistent with Edelberg [2004]. The utility function is defined as in the following:

$$\ln c_1 + \beta \ln c_2 \tag{4.1}$$

where c_i, where $i = 1, 2$ is consumption in the two periods, and β is a discount factor. The budget constraints for the two periods are defined by the following equations:

$$c_1 = L + y_1 \tag{4.2}$$

$$c_2 + L(1 + r) = E[\hat{y_2}] \tag{4.3}$$

where L is a loan, r is the interest rate and y_i, where $i = 1, 2$ is income in the two periods. The budget constraint for period one rules out the possibility that the loan is used for reasons other than consumption, consistently with the data. The risk faced by individuals is income volatility. This is again consistent with the data and the characteristics of the sub-prime credit market (see Alan and Loranth [2013]).

Each individual is characterized by a behavioural score, θ, increasing in their riskiness. The interest rate is an increasing function of individual riskiness. That is, $r = r(\theta)$, with $r'(\theta) > 0$.

Income in period one is observed and not stochastic. Income in period two is stochastic. In particular:

$$E[y_2] = \int_L^H \pi(\theta, y) y \, \mathrm{d}y \tag{4.4}$$

Income spans between 'Low' and 'High', where low income can be zero (unemployment) without loss of generality. The income probabilities are drawn from a joint distribution of income and individual *observed* risk. That is, $\pi = \pi(\theta, y)$. The expected income is found on the marginal distribution of income. The joint probability has the following property:

$$\pi_{\theta,y}(\theta, y) < 0 \tag{4.5}$$

that is, when the individual is riskier it is less likely to jointly observe a high income (and the other way round). This possibility is not ruled out, however. This property has important consequences on the model's predictions.

The individual decides how much to invest in period one. As a consequence, the model firstly concentrates on the optimal loan amount. Then, labour income in period two is realized. After labour income is realized, it is possible to draw conclusions on individual default. The objective function is defined as in the following:

$$\max_L \ln L + y_1 + \beta \ln E[Y_2] - L(1 + r(\theta)) \tag{4.6}$$

The solution of the program is given by:

$$L^* = \frac{E[Y_2] - \beta(1+r)y_1}{(1+r)(1+\beta)} \tag{4.7}$$

and describes the optimal amount of a loan. Although the optimization problem does not admit a corner solution for consumption (due to logarithmic utility), the individual can choose to borrow, to lend or to have a zero amount for L, depending on the numerator of the above equation. In the following we will concentrate on the case in which $E[Y_2] > \beta(1+r)y_1$. This expression implies that if income is on a growing path, this represents an incentive for the individual to borrow more.

It is worth noticing that the optimal loan amount:

- is decreasing in Y_1;
- is increasing in $E[y_2]$;

- is decreasing in the interest rate. This is an important prediction, that will be tested further using regression models;
- is decreasing in θ. This is important, as the model predicts that riskier individuals tend to borrow more.

In period two, individuals default if:

$$c_2 + L^*(1 + r(\theta)) > Y_2^{BM} \tag{4.8}$$

Where BM stands for 'below mean'. These defaults take place for all incomes that are below the expected value of income. That is:

$$Pr.default = \int_L^{Y_2^{BM}} \pi(\theta, y)\,\mathrm{d}y \tag{4.9}$$

The individuals described in so far represent the baseline hazard, that is those individuals who default anyway regardless of the treatment. The following equation

$$c_2 + L^*(1 + r(\theta)) < Y_2^{AM} \tag{4.10}$$

shows that a fraction of people does not default. AM stands for 'above mean'.

The experimental design raises the interest rate in the treatment group by a certain fixed amount r^T. This impacts on the treatment group in the following way:

$$c_2 + L^*(1 + r(\theta) + r^T) > Y_2^{BM} \tag{4.11}$$

that is, those who were supposed to default, default under the treatment as well.

Depending on the other model parameters, those who wouldn't default are now facing the following possibilities:

$$c_2 + L^*(1 + r(\theta) + r^T) < y_2^{HT} \tag{4.12}$$

some of them do not default. HT stands for 'high under treatment'. In the following period these individuals would reduce their credit demand due to the higher interest rate. A fraction of individuals, however, is pushed to default by the treatment:

$$c_2 + L^*(1 + r(\theta) + r^T) > y_2^{LT} \tag{4.13}$$

where LT stands for 'low under treatment'. Under the assumption that riskier individuals are more likely to face low income levels in period two (consistently with equation 4.5), we can conclude that the treatment would push to default borrowers with a poor credit score more relatively more.

4.2.1 Constrained individuals

Suppose that in period 1 individuals face the following budget constraints:

$$c_1 + L_0 = L + y_1 \text{ with probability } p(\theta)$$

$$c_1 = L + y_1 \text{ with probability } (1 - p(\theta))$$

where $p'(\theta) < 0$. That is, individuals are given a negative endowment (for example, a previous loan to repair) with a certain probability that is higher, the riskier the individual.

With the above budget constraints, the optimal loan amount is:

$$\tilde{L^*} = L_0 \frac{\beta}{(1+\beta)} + \frac{E[Y_2] - \beta(1+r)y_1}{(1+r)(1+\beta)} \tag{4.14}$$

with probability $p(\theta)$, and equal to the one in equation 4.7 with probability $(1-p(\theta))$. Equation NUM has a straightforward interpretation. Think of individual i, expecting income in period 2 to be higher than income in period 1. This individual would be a net saver, in absence of a negative endowment. If income in period 1 is high enough, the individual does not need to borrow to repay the negative endowment back. WHowever, if income in period 1 is low, then the individual is forced to borrow, regardless of life-cycle considerations. Notice, moreover, that in the above equation $L_0 \frac{\beta}{(1+\beta)}$ does not depend on the interest rate, implying that an increase in the cost of credit might reduce the optimal loan amount for individuals in this category, but never below a certain threshold. The demand for credit for these individuals should be relatively insensitive to changes in the interest rate. We call these individuals 'constrained borrowers', and we stick to this terminology for the rest of the paper.

All else equal,

$$\tilde{L^*} > L^* \tag{4.15}$$

As a consequence, given the linearity of the problem expressed in equations 4.8 - 4.13, it is clear that individuals who are endowed a negative amount are more likely to default, both in the baseline category and under the treatment.

4.2.2 Outcome variables

The model sketched so far works on the assumptions that individuals live for two periods only, and that any loan to repay is a negative endowment assigned with a random, exogenous probability. Mot only in the database (and in real life) individuals can divide their lifetime in more than two periods, but it is also likely that the loan individuals have to repay is an endogenous decision. Nevertheless, we believe that several of the model predictions can be generalized to more than two periods, and are informative on individual behaviour even under the assumption that past loans are exogenously assigned. This subsection provides further detail about these issues, and motivates the choice of the variables to be introduced in the empirical models.

Consistently with the discussion that follows equation 4.7, our sample contains individuals who are very likely to be net borrowers. As a consequence, we conjecture that an increase in interest rates would unambiguously lower their credit demand.[5] However, as we mentioned above, this statement holds only if the individual is not a 'constrained borrower' (see section 4.2.1). If the individual is constrained, we do not expect credit demand to be particularly responsive to an interest rate increase.

We here draw from Alan and Loranth [2013] to generalize the main results to more than two periods, and to introduce a notation that takes into account the particular characteristics of credit card borrowing. Differently from the model sketched above, the loan is not taken once in a lifetime, but is transferred from period to period on a monthly basis. The notation needs to take this into account.

On a credit card account, an individual does not incur interest charges if they pay the statement balance in full at the specified due date. Therefore the actual monthly credit demand for a credit card user is represented by monthly purchases on credit minus the subsequent payments made

[5]In a basic intertemporal consumption problem, the effect of a change in interest rate has an ambiguous effect on the demand for credit. An interest rate increase would indeed generate both a substitution and income effect. The former is unambiguously negative: higher interest rates lower borrowing [Gambacorta, 2005, page 22], and lead to a reduction in current consumption levels. The latter instead depends on the individual's wealth position. If an individual is a net borrower an increase in interest rates would further reduce current consumption levels. In this case, the overall effect of an interest rate increase is a reduction of credit demand. On the contrary if an individual is a net saver, their income effect would be positive. The total effect of interest rates on consumption cannot be determined *a priori*. If the effect on consumption is positive, a net saver could even increase their credit demand to finance current consumption levels.

toward the outstanding balance. This difference constitutes the monthly addition to the existing credit card debt that accrues interest. Thus, our outcome variable is calculated in this way. We call this indicator net new borrowing, NNB. We define NNB for a generic individual i, at time $t+1$ as:

$$NNB_{it+1} = NT_{it,it+1} - P_{it+1} \tag{4.16}$$

where $NT_{it,it+1}$ is new transactions made by individual i on credit between months t and $t+1$, and P_{it+1} is their payment made toward the outstanding balance at time $t+1$. $NT_{it,it+1}$is interest exempt between period t and $t+1$ whereas if $NT_{it,it+1} - P_{it+1} > 0$, the difference accrues interest charges until paid. Therefore, a positive (negative) value for NNB indicates an increase (decrease) in monthly credit demand.

Assuming that NNB depends on gross interest rates, we expect the unconstrained borrower to reduce NNB when facing an increase in borrowing rates, consistently with equation 4.7. More formally:

$$\frac{\partial NNB_{it+1}}{\partial R_{it+1}} < 0 \tag{4.17}$$

where R_{it+1} is gross interest rate. However, for 'constrained borrowers' we expect very little or no sensitivity to interest rates, consistently with equation 4.14. More specifically:

$$\frac{\partial NNB_{it+1}}{\partial R_{it+1}} = 0 \tag{4.18}$$

Given that now individuals spread their decisions over more than two periods, constrained and unconstrained individuals would as a consequence differ in terms of default behaviour. An unconstrained individual is expected to be insensitive to an increase in interest rate, as they are able to reduce their credit demand. Calling δ the probability of defaulting:

$$\frac{\partial \delta_{it+1}}{\partial R_{it+1}} = 0 \tag{4.19}$$

On the contrary, a 'constrained borrower' would show default sensitivity to an increase in the interest rate.[6] Formally:

$$\frac{\partial \delta_{it+1}}{\partial R_{it+1}} > 0 \tag{4.20}$$

The empirical analyses proposed in the following sections aim at testing these simple theoretical predictions.

[6]Notice that we do not consider the personal cost of default, as the aim of this paper is not to model individual's choice of defaulting. The framework herein developed aims simply at motivating the experimental design and the empirical analyses.

4.3 Data and Experimental Design

As anticipated, the data implemented to test the above theoretical predictions come from a unique database made available by a private lender. The company routinely operates randomized trials to measure the clients' sensitivity to the interest rate. The analyses performed make use of a peculiar experimental design by the lender in November 2007. The section is organized as follows. Section 4.3.1 provides a general description of the lender and the data. Section 4.3.2 describes the experimental design. Section 4.3.3 describes the data and provides summary statistics. Section 4.3.4 provides tests of internal validity.

4.3.1 The lender

Our lender operates in the sub-prime segment of the UK market.[7] Given the limited number of non-standard credit card issuers in the UK, for confidentiality reasons we do not divulge the exact market share of the lender. However, for descriptive purposes, it can be said that the lender plays a major role in the UK subprime market. It offers several credit card products with characteristics typically observed in this niche, such as high interest rates and low credit limits [Einav et al., 2012].

The lender has routinely been performing randomized interest rate experiments on sub-samples of its clients since 2006. These experiments have the main target of monitoring clients' sensitivity to the interest rate. Each experiment lasts around 3-6 months and the lender generally initiates a new experiment immediately after the previous one. Interest rate changes are permanent until the next change is implemented. The proportion of individuals allocated to control groups has become increasingly smaller with respect to earlier rounds. All interest rate experiments have been designed based on ex-ante determined blocks. How the lender identified the blocks for the present experiment will be explained in greater detail in the following section.

4.3.2 The experimental design

Regarding the experiment herein discussed, the lender carried out the randomization as a block design where randomly selected samples of individuals were assigned to 'cells', as the lender names them. For the ease of exposition we will use their term throughout the paper. Cells were in turn defined by the interaction between utilization rates (month end balance divided

[7]For a definition of sub-prime credit market in the United Kingdom see PFRC - Bristol [2013], BIS [2013]

by credit limit) and internally developed credit scores ('behavioural scores').[8] The lender chose this particular design mainly to structure the best risk pricing practice via proper segmentation of the accounts. However, block designs have also the advantage that the ex-ante conditioning on observable characteristics yields more precise estimations of average treatment effects, and are indeed commonly used in field experimental designs.

More in detail, for each variable (utilization rate and internally developed credit score), the lender identified three 'bands': low, medium and high. As a consequence, the interaction between the two variables identifies 9 cells, as the following figure describes.

Utilization Rate	Behavioural Score: Low		Mid		High	
High	Cell 1	4599	Cell 4	12448	Cell 7	17555
	Control	280	Control	788	Control	1128
	Treatment	4319	Treatment	11660	Treatment	16427
Mid	Cell 2	281	Cell 5	2317	Cell 8	8123
	Control	0	Control	0	Control	518
	Treatment	281	Treatment	2317	Treatment	7605
Low	Cell 3	137	Cell 6	793	Cell 9	5096
	Control	0	Control	0	Control	323
	Treatment	137	Treatment	793	Treatment	4773

FIGURE 4.1: Experimental Design

It is worth noticing that in each cell the lender allocated a small fraction of costumers to the control group: roughly 6.5% in each of them. Moreover, the lender did not allocate any control groups in cells 2,3,5 and 6, making them unusable for our study. These cells will not be considered in the analyses of the paper from now onwards, unless mentioned otherwise. In all cells, the treatment consisted in a 5 percentage points increase in the clients' interest rate. This increase applied to all treated individuals regardless of their pre-treatment interest rates. The pre-treatment interest rate distribution shows considerable volatility: from to minimum value of 20.95% to a maximum value of 34.92%, with a median value of 29.9%.

[8]Notice that the score decreases in individual risk, such that high scores correspond to low risk. Internally developed credit scoring systems are general practice for subprime lenders. In fact, they heavily invest to develop precise default prediction algorithms using multivariate analysis. Although the exact features of the lender's scoring system was not disclosed, we were informed that it is a continuously updated multivariate probit type system.

4.3.3 Data and descriptive statistics

The data set contains rich and detailed information on individual transactions but offers, like many propriety transactions data, limited information on demographics. It comprises transactions including purchases, payments, cash withdrawals, statement balance and interest charges. Moreover, it provides information on the characteristics of the contract, such as credit limit, presence of direct debit payment authorization and interest rates. We have limited demographic information such as income, age, employment and marital status, which were reported by the client at the application stage. Unfortunately, we do not have information on individuals' other credit commitments such as mortgages and other consumer loans, or on their collateral.

The experimental sample was not chosen from the lender's full clientele base. Accounts that are flagged for reasons such as default, several months of delinquency or inactivity were excluded before the selection of the sample. Furthermore, the lender excluded individuals who had been with the lender for less than seven months at the time of the design (November 2007). Table 4.1 presents the characteristics of the individuals in the sample. Values are calculated in the month in which individuals were assigned to treatment and control groups (November 2007).

TABLE 4.1: Descriptive statistics

	Mean	Median	Std. Dev
Utilization rate (%)	0.79	0.95	0.33
Statement balance (£)	791.88	678.75	612.92
debt2	753.57	640.64	609.38
New retail transactions (£)	67.23	0.00	163.80
New cash withdrawal (£)	13.85	0.00	61.77
Credit limit (£)	1105.30	900.00	767.10
Interest rate (Nov.2007)	0.31	0.30	0.02
Income	18091.64	15600.00	36589.84
Age	38.43	37.00	11.59
Married (%)	0.43	0.00	0.50
Employed (%)	0.65	1.00	0.48
Self employed (%)	0.05	0.00	0.23
Home owner (%)	0.21	0.00	0.40
No other cards (%)	0.43	0.00	0.50
Observations	47821		

The values presented in this table highlights the distinct characteristics of sub-prime borrowers. The average credit limit is £1,127, which is significantly lower than the credit limits typically found in the standard market . Individuals are charged fairly high interest rates (33.24% pa on average), compared to the standard market (approximately 15-18% pa on average) [Data Monitor UK, 2008]. Their average utilization rate (ratio of the statement balance to credit limit) is higher than the one typically observed in the prime market (79% vs 34%) [Data Monitor UK, 2008].

As already mentioned, the paper assesses whether the treatment has an effect on credit demand and individual default, and analyzes whether this effect is heterogeneous across cells. Credit demand and default are as a consequence the main dependent variables of our empirical models. Credit demand is defined in section 4.2, and calculated as therein described. Default is defined as in the following. In our data, if an individual is delinquent for four consecutive months of delinquency, the interest charges (and late fees) disappear from their records. If two or more months of delinquency follow this four-month delinquency period, the individual's account gets charged off and the outstanding debt is removed from the data base. We define individuals as defaulters under these circumstances, and therefore we observe individual defaults six month after the beginning of the treatment (in June 2008).

4.3.4 Internal validity tests

Our lender conducted this experiment as part of its risk based pricing practice so we believe that they have every incentive to execute it correctly. However, it is important to verify that the randomization was carried out properly to ensure the internal validity of our results.

Table 4.2 presents balancing results for each cell as at November 2007.

TABLE 4.2: Internal Validity Checks.

	Cell 1 Mean(T) - Mean(C)	Cell 4 Mean(T) - Mean(C)	Cell 7 Mean(T) - Mean(C)	Cell 8 Mean(T) - Mean(C)	Cell 9 Mean(T) - Mean(C)
Utilization rate (%)	1.06 — 1.04	0.97 — 0.98	0.94 — 0.94	0.50 — 0.51	0.04 — 0.05
	(0.23)	(0.01)	(0.73)	(0.91)	(0.49)
Statement balance (£)	837.28 — 843.92	766.05 — 759.87	1033.22 — 1028.24	699.21 — 694.24	192.37 — 160.34
	(0.84)	(0.75)	(0.81)	(0.83)	(0.18)
Behavioural Score	586.52 — 585.61	504.32 — 481.50	636.27 — 633.42	636.22 — 622.19	635.13 — 629.41
	(0.91)	(0.03)	(0.68)	(0.19)	(0.69)
debt2	812.40 — 821.39	744.39 — 739.38	1005.35 — 1002.02	635.52 — 635.69	47.60 — 58.74
	(0.78)	(0.79)	(0.87)	(0.99)	(0.41)
New retail transactions (£)	13.68 — 9.03	43.36 — 39.57	80.71 — 74.02	134.00 — 131.49	61.41 — 57.76
	(0.19)	(0.40)	(0.22)	(0.79)	(0.66)
New cash withdrawal (£)	3.96 — 3.24	14.64 — 12.47	17.28 — 16.59	19.33 — 20.37	7.73 — 5.75
	(0.59)	(0.28)	(0.76)	(0.73)	(0.38)
Credit limit (£)	798.18 — 811.12	800.22 — 788.66	1131.45 — 1121.99	1461.20 — 1447.30	1529.57 — 1537.48
	(0.69)	(0.55)	(0.67)	(0.73)	(0.88)
Interest rate (Nov.2007)	0.31 — 0.31	0.31 — 0.31	0.31 — 0.31	0.31 — 0.31	0.30 — 0.30
	(0.58)	(0.83)	(0.20)	(0.11)	(0.62)
Income	17866.19 — 17672.64	17185.75 — 17597.75	17251.63 — 17900.43	17560.80 — 19335.33	18497.04 — 18745.22
	(0.84)	(0.33)	(0.07)	(0.09)	(0.74)
Age	36.61 — 36.67	37.42 — 37.48	39.27 — 39.66	39.23 — 39.02	37.34 — 37.23
	(0.93)	(0.88)	(0.28)	(0.69)	(0.88)
Net New Borrowing (£)	-11.21 — -13.51	21.67 — 19.06	52.83 — 47.79	70.32 — 72.90	-96.56 — -52.73
	(0.63)	(0.54)	(0.35)	(0.80)	(0.04)
Observations	4599	12448	17555	8123	5096

Mean values for treatment (T) and control (C). Estimation is performed in the month of randomization (November 2007). P-values for unpaired t-tests (Welch) are in parentheses.

We do not observe any statistically significant mean differences between treatment and control groups in any cell. All tests are performed at the 5% level, at the month of November 2007.

The test are performed without imposing equal variances in the treatment and the control group (Welsh test). Imposing equal variances (Bonferroni test) does not change the results. Distribution equality tests using Kolmogorov-Simirnov and K-Wallis tests were also performed, and could not detect any statistically significant difference between the treatment and the control groups.

Randomization is not the only source of potential concern for the validity of an experimental design. Sample attrition, for example, can be of particular harm if caused by the treatment. However, this attrition is the interest of this paper, i.e. we are interested in whether the treatment (a 5 percentage point increase in interest rates) generates individual default. The database does not contain any individual drop-outs within the sample period and we do not face attrition in the traditional sense. Attrition does not represent a major concern for the validity of our results.

4.3.5 External validity

Concerns might arise with respect to the external validity of the empirical results, as the experiment takes place exactly during the main 2007-2008 financial crisis events. These very same events might either have influenced the outcomes of the research, or have lead to results whose importance is circumscribed to this particular period. We will briefly explain why in both cases the financial crisis does not represent a source of concern for the external validity of the results.

With respect to the first criticisms, it is important to recollect that the experiment has been run on a sample of individuals who had been clients of the lender for more than 7 months at the time of the experimental design (November 2007). As the first events related to the financial crisis date back to August 2007, we can exclude that the sample of costumers has been relying on subprime credit because of the adverse economic downturn. Moreover, the experiment actually takes place during the first six months of 2008, with default observed during the month of June. This happened before the main disruptive event of the 2007-2008 crisis (the bankruptcy of Lehman Brothers, taking place on September 15^{th} 2008), leading us to think that the outcome of the experiment has not been significantly influenced by the financial crisis.

Having said so, it is left to understand whether the empirical results are of general interest and still relevant in the post-crisis. The answer is yes. As Personal Finance Research Centre [2009] and Personal Finance Research Centre [2011] report, the global financial crisis has left a scar on the personal finances of UK households. If in 2009 the worsening of UK household finance conditions was still not so evident [Personal Finance Research Centre, 2009], in 2011 we observe a more pronounced change in the financial conditions and habits of UK households, with an increasing propensity to use unsecured debt and to rely on subprime credit to cope with everyday expenses. The recent debate with respect to the introduction of a price-cap on high cost credit [Financial Conduct Authority, 2014, Personal Finance Research Centre, 2014] testifies the growing need to understand more on subprime credit behaviour, strengthening the relevance of our results.

4.4 Results

This section discusses the empirical results.

4.4.1 Empirical Model

To estimate the average treatment effects of default probability, we run logit regressions. Estimations are performed cell by cell, during the month of June 2008. The model is described by the following equation:

$$Pr(Default_{ij} = 1) = \Lambda(\alpha_0 + \alpha_1 T_{ij} + \alpha_2 X_{ij} + \varepsilon_{ij}) \tag{4.21}$$

where Λ is the logistic density function; $Default_{ij} = 1$ is a dummy variable equal to 1 if the individual i in cell j is a defaulter, and equal to zero otherwise (see section 4.3.4); T is a dummy variable which equals 1 if individual i in cell j is in the treatment group and zero otherwise; X_{ij} is a vector of observables for individual i in cell j that are potentially predictive of the default behavior. The vector contains their pre-treatment interest rate, income, age, marital stuatus and employment status. We are interested in estimating the treatment effect on the default probability. Since the randomization ensures that treatment status T is orthogonal to all observable and unobservable individual characteristics in a given cell, α_1 provides an unbiased estimate of the treatment effect on default probability. Moreover, since the covariates we use in our regressions are independent on the treatment status (see table 4.2), including them in the analysis is expected to improve the precision of the estimated treatment effect to the extent that they are predictive of the outcome (default).

Similarly, to estimate the average treatment effects of credit demand, we run OLS regressions. Estimations are performed cell by cell, over the March - June 2008 period. Standard errors are clustered by id. The model is described by the following equation:

$$NNB_{ij} = \beta_0 + \beta_1 T_{ij} + \beta_2 X_{ij} + e_{ij} \tag{4.22}$$

where NNB_{ij} is net new borrowing, and the remaining variables are defined as above. Again, since the randomization ensures that T is uncorrelated with unobservables contained in e_{ij} and all other observable characteristics, β_1 provides an unbiased estimate of the treatment effect on the average credit demand for the March - June 2008 time span.

4.4.2 Main empirical results

Before discussing the main results, we display evidence that without an exogenous variation in the interest rate we cannot identify its causal effect on individual default. To illustrate this point we estimate a logit model of individual default on their risk score, utilization rate and post-treatment interest rate. The results are contained in table 4.3.

TABLE 4.3: Logit Model. Dependent Variable: Default.

	(1)
Behavioural Score	-0.00015***
	(0.00)
Utilization rate (%)	0.11969***
	(0.01)
Interest Rate (June)	-0.09456**
	(0.04)
Income	-0.00000
	(0.00)
Age	-0.00017*
	(0.00)
Employment status	yes
Marital status	yes
Observations	48292

The estimations is performed on a cross section of clients observed in the month of June 2008. The reported coefficient are marginal effects. Inference is based on a robust variance covariance matrix. An individual is considered a defaulter if either charged-off or after six consecutive months of delinquency. Behavioural Score is an internal client risk assessment operated by the lender. Utilization rate is the ratio between statement balance and credit limit. Interest Rate (June) is the post-treatment interest rate applied to the client. Income is individual income, measured in pounds. Age is individual age, measured in years. Employment and marital status are a set of binary indicators pertaining to these categories. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The dependent variable is the probability of defaulting during the month of June 2008, as defined in the previous section. In absence of experimental data, this is the regression one would normally conduct. The estimation reveals how endogeneity influences the coefficient of the independent variables. As predicted, a higher behavioral score is associated with a lower default rate, whilst a higher utilization rate leads to higher default rates (Agarwal et al. [2010]). However, not taking into account that interest rates are endogenous to borrower characteristics leads to a misleading result: the correlation between interest rate and default probability is indeed negative. With this in mind, we can comment on our main results where we have complete information on the counterfactual thanks to block randomization.

Table 4.4 presents the estimated treatment effect of a 5 percentage point increase in the interest rate on the probablity of defaulting and on credit demand. Both models are estimated cell by cell.

TABLE 4.4: Regression Models. Dependent Variables: Defualt ; Net New Borrowing.

	Cell 1		Cell 4		Cell 7		Cell 8		Cell 9	
	Default	NNB	Default	NNB	Default	NNB	Default	NNB	Default	NNB
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. T. E.	0.013	-2.735	0.022**	-1.867	-0.003	-9.674***	0.007	-1.453	-0.003	-1.269
	(0.03)	(3.89)	(0.01)	(2.30)	(0.01)	(3.46)	(0.01)	(7.03)	(0.00)	(9.09)
Interest rate (Nov.2007)	-0.325	-5.161	-0.044	18.302	-0.080	36.884	-0.008	-4.311	0.050	-96.621
	(0.25)	(40.51)	(0.09)	(29.46)	(0.06)	(36.48)	(0.06)	(81.79)	(0.07)	(119.91)
Income	-0.000	0.000*	-0.000	0.000	-0.000	0.000***	-0.000	0.000	-0.000	0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age	0.001	-0.350***	-0.001***	-0.365***	0.000	-0.538***	-0.000	-0.371**	-0.000	-0.170
	(0.00)	(0.09)	(0.00)	(0.07)	(0.00)	(0.08)	(0.00)	(0.18)	(0.00)	(0.26)
Baseline Value	0.200***	-24.323***	0.049***	-21.319***	0.035***	-16.866***	0.006*	13.197*	0.009*	49.435***
	(0.02)	(3.74)	(0.01)	(2.20)	(0.01)	(3.36)	(0.00)	(6.81)	(0.01)	(8.71)
Employment Status	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Marital Status	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4591	4591	12442	12442	17551	17551	7811	7811	4594	4594
M.D.E. (μ_2)	0.26	-35.52	0.07	-27.07	0.05	-25.19	0.142	-4.32	0.02	23.44

Odd columns contain logit models. All estimations are performed on a cross section of clients observed in the month on June 2008. The reported coefficient are marginal effects. Inference is based on a robust variance covariance matrix. An individual is considered a defaulter if either charged-off or after six consecutive months of delinquency. Even columns contain OLS models. All estimations are performed on a panel of clients observed between the months of April and June 2008. Inference is based on a variance covariance matrix clustered by individuals. Net new borrowing is defined as the monthly increase in credit card debt for retail purposes minus the monthly payment and interest rate charges. In all regression, Interest Rate (2007) is the pre-treatment interest rate applied to the client. Income is individual income, measured in pounds. Age is individual age, measured in years. Employment and marital status are a set of binary indicators pertaining to these categories. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. M.D.E. stands for minimum detectable effects. M.D.E. are calculated using sample estimations, and setting $\alpha = 0.05$ and $1 - \beta = 0.80$. Reported, is the value the mean would have to take under the treatment for a significant treatment effect to be detected at the 5% level.

Cells 8 (low risk, medium utilization rate) and 9 (low risk, low utilization rate) show a very similar behaviour to each other. In both cells individuals in the control group have very low default rates, increase their credit demand over the March - June 2008 period, and show no treatment effect both on default and net new borrowing. The cells do not show particular power. However, in both cells and for both variables the means under the treatment and the control are fairly similar. We believe that even if the cells had more power, there would not be a sizable effect to detect.

Cell 7 presents more interesting results. Individuals in this cell show a non-negligible baseline default rate (3.5%) and overall reduce their credit demand in the control group. The cell is characterized by a sizable negative treatment effect in credit demand ($-9.674, p-value : 0.00$). This implies that treated individuals reduce their credit demand 57% more, on average. Cell 7 displays instead a close-to-zero (and not significant) treatment effect on default. The behaviour of cell 7 individuals is consistent with the result of equations (2) and (4) (section 4.2). Low risk individuals do not seem to be borrowing constrained, as they reduce their credit demand when facing an increase in their interest rate, but show no interest rate sensitivity with respect to default. The cell is numerous enough to have power: it has the power to detect both the difference in default probability, and the difference in credit demand.

Cell 4 behaves consistently with the theoretical predictions. Similarly to cell 7, individuals in this cell show a sizable baseline default rate (4.9%) and overall reduce their credit demand in the control group (-21 pounds over the March-June 2008 period). The treatment effects are characterized by an opposite behaviour, though. We observe a positive and significant treatment effect on the probability of default: 2.2%. This corresponds to a 45% increase with respect to the baseline value. Moreover, we observe an overall small and non-significant treatment effect on credit demand. The predicted change is 8% of the baseline value. However, the coefficient is not significant despite the cell has the power to detect an effect (see the bottom of the table). Moreover, such a coefficient would represent a reduction in credit demand of 45p per month. We consider this amount small enough to conclude that individuals in these cell show no sensitivity of credit demand with respect to an increase in the interest rate. Consistently with equations (3) and (5) we conclude that individuals in cell 4 are on average borrowing constrained. Power analysis partially supports our results: the cell has indeed the power to detect the treatment effect on default. The credit demand effect is instead negligible in itself, and does not need further comments.

The comments referring to Cell 4 apply to Cell 1 as well. Individuals in this cell show a huge baseline default rate (20%), comparable to the findings of non-academic reports on the topic [OFT, 2010]. Cell 1 individuals overall reduce their credit demand in the control group (-24.3 pounds over the March-June 2008 period). We observe a positive and sizable treatment effect on the probability of default, but this effect is not significant: 1.3%. This implies that individuals in the treatment group default on average 6.5% more than individuals in the control group. Moreover, we observe an overall small and non-significant treatment effect on credit demand. The predicted change is 10% of the baseline value. Nowever, the coefficient represents, in absolute terms, a reduction in credit demand of 68p per month. We consider this value small enough to claim that these individuals do not really show a reduction in credit demand because of the treatment.

Although individuals in cell 1 behaves consistently with the presence of borrowing constraints, we are not able to identify a significant treatment effect on default probability. We have a couple of explanations for this finding. Firstly, cell 1 is overall more heterogeneous than the other cells. For example, new clients are given low behavioural scores just because they lack credit history with the lender. This does not make them intrinsically riskier, though. As a consequence, the treatment effect is likely to be less sizable than it would be in a more homogeneous cell. Needless to say, a smaller treatment effect is by definition more difficult to detect. Moreover, the sample size of cell 1 is seinsibly lower than than the ones of Cells 4 and 7. It is 36% of cell 4, and 26% of cell 7. Power analyses show that cell 1 simply does not have the power to detect an effect that we argue would be significant otherwise.[9]

4.4.3 Interest rate sensitivity of returns

Lenders face a trade-off when setting their interest rates. Higher interest rates can increase revenues, but they can also reduce the quantity demanded and increase the lender's costs by causing more individuals to default. In this section we assess whether the increase in the interest rate associated with the treatment resulted in an overall profitable intervention for the lender. The treatment caused indeed some individuals to reduce their credit demand (cell 7), and some to default (cell 4, and cell 1 with some caveats : see the previous section).

To investigate the above question, we calculate the interest charges of each individual in the sample, and we perform an OLS regression of interest charges on treatment and the controls described above. Regressions are performed cell by cell. Similarly to what we did with NNB, the analyses are performed on the March - June 2008 period. Inference is based on a cluster variance-covariance matrix.

Table 4.5 contains the regression results.

As columns (1) to (5) suggest, the intervention resulted profitable for all cells apart from cell 9.[10] Cell 8, where we do not observe any sizeable treatment effect on default or credit demand, is the cell that show the higher profitability: the treatment effect represent almost 17% of the baseline interest rate charges. Cells 1 and 7 follow, with a treatment effect close to 10% of the baseline interest rate charges. Despite the significant reduction in credit demand, and the (non-significant, though) increase in defaults, in these cell the treatment managed to increase the interest rate charges. In cell 4 the treatment effect is relatively less profitable: 5% of the baseline interest charges. This result can be attributed to the high number of defaulters generated by the treatment. A client who reduces their credit demand can indeed still generate

[9]Although cells 2, 3, 5 and 6 do not contain a control group, it is worth checking whether they present comparable default rates with respect to the other cells. Cell 2 is the riskiest, and present indeed the highest default rate (11%). The other cell behave accordingly, with the lowest default rate found in the least risky cells (4% in cells 5 and 6). Results are displayed in table 4.8 in the appendix.

[10]We are considering the interest charges defaulters do not pay as a the only cost the lender is bearing for filing. We do not observe actual filing costs in the database, and for simplicity we are assuming they are equal to zero. These profits should be read as an upper bound.

TABLE 4.5: OLS regressions. Dependent Variable: Interest Rate Charges.

	Cell 1	Cell 4	Cell 7	Cell 8	Cell 9
	(1)	(2)	(3)	(4)	(5)
A. T. E.	6.203**	4.015**	9.718***	13.278***	-2.006
	(2.96)	(1.85)	(1.99)	(2.74)	(3.11)
Interest rate (Nov.2007)	31.112	-45.111**	-110.157***	132.676***	128.874***
	(35.32)	(22.13)	(24.18)	(37.65)	(35.34)
Income	0.000	0.000	0.000	-0.000***	-0.000*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age	0.275***	0.764***	0.857***	1.014***	0.233***
	(0.09)	(0.05)	(0.06)	(0.08)	(0.08)
Baseline Value	60.091***	73.728***	96.888***	81.627***	36.741***
	(3.06)	(1.83)	(1.96)	(2.70)	(3.16)
Employment Status	Yes	Yes	Yes	Yes	Yes
Marital Status	Yes	Yes	Yes	Yes	Yes
Observations	13794	37339	52665	24369	15284

All columns contain OLS models. All estimations are performed on a panel of clients observed between the months of April and June 2008. Inference is based on a variance covariance matrix clustered by individual. Interest rate charges are as the monthly sum of interest monthly interest rate charges from January 2008 till the month in which the individual is observed. In all regression, Interest Rate (2007) is the pre-treatment interest rate applied to the client. Income is individual income, measured in pounds. Age is individual age, measured in years. Employment and marital status are a set of binary indicators pertaining to these categories. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

revenue. However, a client who defaults does not generate revenues by definition. Nevertheless, the intervention resulted profitable in cell 4 as well.

This exercise shows that despite the treatment represents a reasonably big increase in the interest rate (5%), and despite individuals show a certain credit demand and default sensitivity to the interest rate increase, the intervention managed to increase interest rate charges. The price increase counterbalanced the quantity reduction in cell 7, generating higher revenues. The higher revenues generated in cells 1 and 4 counterbalanced instead the filing costs.

It is worth mentioning that the exercise herein described discusses profitability and the sensitivity of returns for the existing pool of costumers only. The purpose of the company is to understand how sensitive existing costumers are, and how an increase in the interest rate would affect profitability conditional on the clientele base already collected. This exercise is not informative on how and increase in the interest rate would affect the ability of the lender to attract new costumers, or to retain the existing ones in the long run.[11]

[11]Coeteris paribus, higher interest rates would make the credit card product less attractive. With respect to the ability to attract new costumers, economic theory can be informative to predict two different (and not mutually exclusive) outcomes. The first is straightforward. All else equal, if the price of obtaining credit goes up, the lender will be able to attract less costumers. The second, less intuitive, is that the worse credit conditions would increase asymmetric information, sorting out bad credit (adverse selection) and creating incentive not to repay (moral hazard). While the first prediction does not allow us to predict whether returns will go up or down (as this depends on the elasticity of credit demand), the second would surely decrease returns, as higher adverse selection and moral hazard would lead to higher default probabilities. With respect to the long run capacity to retaining existing costumers, this depends on the characteristics of the pool of clients. Constrained clients, with

4.4.4 Robustness check: transitions

At the time of the experimental design (November 2007) the behavioural scores of individuals in different risk categories (low, medium, high) for a given level of utilization rate do not overlap across cells. The same happens for utilization rates, given a certain risk category. After the beginning of the experiment, however, behavioural scores and utilization rates are allowed to change. As a matter of fact they start overlapping across cells. The following graph describes the November to June transition of behavioural scores.

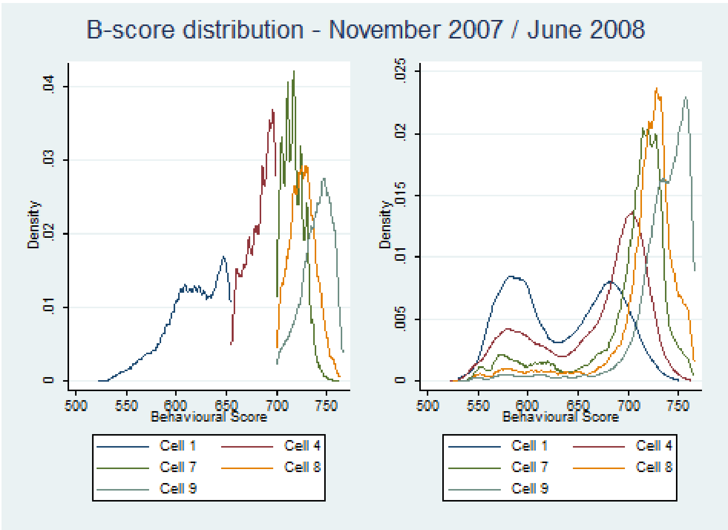

FIGURE 4.2: Bscores by cell - before and after the treatment

As a robustness check, we analyze whether the treatment caused individuals to worsen their situation. The rationale of this analysis is to investigate whether individuals who are pushed to a 'worse cell' by the treatment are likely to be our defaulters. To perform this analysis we reallocated individuals to cells in June 2008, making use of the behavioural score and utilization rate thresholds implemented in November 2007. We call 'stayers' individuals who, in June 2008, are allocated to the same cell from which they started in November 2007. We call 'worse-off individuals' clients who make a transition to a lower-ranked cell (say, from 7 to 5, or from 4 to 3). We call 'better-off individuals' clients who make a transition to a better-ranked cell (say, from 3 to 6, or from 1 to 3).[12]

no access to other sources of credit, will remain clients, if in need for credit to cope with their everyday activities. Unconstrained clients, with access to other sources of credit, would be more likely to change lender, all else equal. See also section 4.4.6 for a discussion on this.

[12]It goes without saying that an individual with a lower behavioural score did not improve their position or their financial well-being. We argue that having a higher utilization rate implies being more financially exposed. As a consequence we consider an individual who stays at the same level of risk, but is more financially exposed, worse-off, given the particular nature of the sub-prime credit market.

The estimated regression model is the following:

$$Pr(Transitionfromcellj_i) = \Lambda(\gamma_0 + \gamma_1 Treated_i + \gamma_2 X_{ij} + \epsilon_i) \tag{4.23}$$

where the transition, as anticipated, can be: staying, going to a worse cell, going to a better cell; and the rest of the equation has been defined above (section 4.4.1). The sample is restricted to individuals being charged the *median* pre-treatment interest rate (0.295). Although this restriction reduces the sample size, it also reduces the heterogeneity of the sample, and actually helps improving power. More details about power are contained in the following. Moreover, the vast majority of the individuals in each cell are charged the median interest rate, so the analysis is a relevant description of cell behaviour. Table 4.6 contains the regression results.

TABLE 4.6: Logit Models.

Panel 1	Dependent Variable: staying in a given cell.				
From:	Cell 1	Cell 4	Cell 7	Cell 8	Cell 9
A. T. E.	0.091**	-0.001	0.008	-0.040	0.067*
	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)
Baseline Probability	0.317***	0.275***	0.501***	0.354***	0.498***
	(0.04)	(0.02)	(0.02)	(0.03)	(0.03)
Controls	x	x	x	x	x
Observations	2409	6982	10061	4813	3167
Panel 2	**Dependent Variable: moving to a worse cell.**				
From	Cell 1	Cell 4	Cell 7	Cell 8	Cell 9
A. T. E.	-	0.037*	-0.011	0.044	-0.072**
	-	(0.02)	(0.02)	(0.03)	(0.03)
Baseline Probability	-	0.266***	0.291***	0.437***	0.455***
	-	(0.02)	(0.02)	(0.03)	(0.03)
Controls	-	x	x	x	x
Observations	-	6982	10061	4813	3167
Panel 3	**Dependent Variable: moving to a better cell.**				
From	Cell 1	Cell 4	Cell 7	Cell 8	Cell 9
A. T. E.	-0.091**	-0.035	0.004	-0.002	-
	(0.04)	(0.02)	(0.02)	(0.02)	-
Baseline Probability	0.683***	0.459***	0.208***	0.209***	-
	(0.04)	(0.02)	(0.02)	(0.02)	-
Controls	x	x	x	x	-
Observations	2409	6982	10061	4813	-

Panel 1 of table 4.6 describes the staying-on probability. Consistently with a borrowing constrains explanation, individuals in cell 1 are kept in their cell by the treatment. More in detail, the treatment increase the probability of staying in cell 1 by 9%. This is a sizeable effect: 29% of the baseline probability. Power analyses show that this mean difference can be detected with a power of 0.7, which we consider a reasonably high number.

Panel 2 of table 4.6 describes the probability of moving to a worse cell. Consistently, cell 4 displays a positive and significant treatment effect. 3.7% of the individuals are puhed to an inferior cell by default. This represents 13% of the baseline value. Such an effect would be safely detected even with a power of 90%. However, two questions remain open at this point. Firstly we need to understand whether these individuals are borrowing constrained. Secondly, we need to understand whether these individuals are our defaulters. To answer the fist question, we claim that a borrowing constrained individual in cell 4 would who worsens their situation would move to cell 1, but not to cells 2 or 3. This because they should not be able to reduce their credit demand. Further analyses demonstrate that this is the case: we find a positive and significant treatment effect on the probability of moving from cell 4 to cell 1 ($0.04, p - value : 0.07$), a negative and not-significant treatment effect on the probability of moving from cell 4 to cell 2 ($-0.002, p - value : 0.25$), and no treated individuals moving from cell 4 to cell 3. To answer the second question, we create an interaction variable between the default dummy and the 'worsening' dummy (from cell 4), and we perform a logistic regression on treatment. We find that the treatment has a positive and significant effect on the probability of worsening from cell 4 *and* defaulting ($0.016, p - value : 0.06$).

Panel 3 simply shows that treated individuals in cell 1 are less likely to improve their situation: the cell 1 stayers of panel 1. This result does not need further comments.

4.4.5 Robustness check: duration models

Gross and Souleles [2002] analyze data that are similar, in their structure, to the ones herein implemented. They have monthly credit card transaction data, in which they observe delinquency, default, and a relatively limited number of demographics. In their contribution, they estimate duration models of default. For comparability, and as a robustness check, we perform a similar analysis. It is worth noticing, however, that this cannot be our main analysis: we do not observe individuals longer than six months after the implementation of the experiment (till July 2008). Default however cannot take place before the sixth month of delinquency. As a consequence, the duration models we estimate do not contain such a different information with respect to the binary choice models estimated above. It is anyway relevant to include such analyses, for comparability with the related literature and to check the robustness of the results obtained.

Table 4.7 contains the main results. Odd columns contain parametric proportional hazard models with an exponential specification. Even columns contain non parametric proportional hazard models (Cox models). The period considered is February - July 2008.

Table 4.7: Duration models (robustness check).

	Cell 1		Cell 4		Cell 7		Cell 8		Cell 9	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. T. E.	1.034	1.034	1.208*	1.210*	0.869	0.868	1.419	1.420	0.656	0.656
	(0.10)	(0.10)	(0.13)	(0.13)	(0.10)	(0.10)	(0.48)	(0.49)	(0.26)	(0.26)
Income (£)	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age (years)	1.000	1.000	0.997	0.997	0.996	0.996	0.983**	0.983**	0.985	0.985
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)
Employment Status	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Marital Status	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4591	4591	12442	12442	17551	17551	7811	7811	4594	4594

Odd columns contain parametric models, and make use of an exponential survival distribution. Even columns contain non-parametric models (Cox models). All estimations are performed on a panel of clients observed between the months of February and July 2008. The reported numbers are exponentiated coefficients. They represent the change in the hazard rate of defaulting associated with a marginal increase in the regressor. An individual is considered a defaulter if either charged-off or after six consecutive months of delinquency. In all regression, Interest Rate (2007) is the pre-treatment interest rate applied to the client. Income is individual income, measured in pounds. Age is individual age, measured in years. Employment and marital status are a set of binary indicators pertaining to these categories. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The results are consistent with the one commented in section 4.3.2: we find a significant treatment effect in cell 4. The treatment increases the hazard rate of defaulting in June or July 2008 by 20% (columns (3) and (4)). The result is consistent to parametric estimation using alternative distributions (Weibull).

The analyses commented above present two drawbacks. Firstly, PH models assume that individual characteristics do not influence latent survival time. This assumption is restrictive, however in our case it could represent a good description of the data, as individuals cannot default before the sixth month of delinquency. Estimation of an Accelerated Failure Time Weibull model does not change the main results.

The second drawback is that monthly data cannot be treated as continuous, in theory. Gross and Souleles [2002] estimate indeed dynamic probit models to model the time dependence of the defaulting event. We estimate a probit model of default with time dummies and clustered standard errors on a February - July 2008 sample, and we find consistent results: individuals in cell 4 default more because of the treatment, and the result is robust to the introduction of the month dummies ($0.01, p - value : 0.06$). Individuals in the other cells do not show significant treatment effects.

4.4.6 Access to other sources of credit

If individuals have access to alternative sources of credit, this can influence the baseline theoretical predictions linked to an increase in the cost of borrowing. If an individual has access to other credit opportunities, two motives are indeed present:

- a *substitution* motive: if the price of credit in a given alternative goes up, the individual can choose to increase their credit demand in a cheaper borrowing option;
- a *hedging* motive: as an increase in the cost of credit represents an increase in default probability, an individual might use less risky options to hedge their risk. Potentially, they can reduce their credit demand from the risky line of credit by borrowing money from a safer lender.

The first motive leads to the conclusion that an individual with no access to other sources of credit is less able to reduce their credit demand. The second motive leads to the conclusion that an individual with no access to other sources of credit is more prone to default.

As we are able to observe whether the clients have access to their credit card, the data allow us to test the above predictions, at the same time checking the robustness of the experimental design. The treatment is indeed orthogonal to all individual characteristics, access to alternative sources of credit included. The introduction of this additional indicator in the reduced form models should not affect the magnitude and the significance of the treatment coefficient.

The following table contains the regression results:

Consistently with what mentioned above:

- the significance, the sign and the magnitude of the treatment effects remain unchanged with respect to the previous specification (see table 4.4 for a comparison);
- individuals with no other cards are less able to reduce their credit demand (consistently with the *substitution* motive);
- individuals with no other cards are slightly more prone to default (consistently with the *hedging* motive).

Unfortunately, our data is overall limited with respect to individual characteristics, and in particular on their wealth. As a consequence, providing more involved evidence on the topic is objectively not feasible. Nevertheless, this robustness check shows that the overall findings of the paper, and the experimental design, are robust to controlling for additional sources of credit. Moreover, the additional regressor behaves according to the theoretical predictions.

Table 4.8: Regression Models. Dependent Variables: Defualt ; Net New Borrowing.

	Cell 1		Cell 4		Cell 7		Cell 8		Cell 9	
	Default	NNB	Default	NNB	Default	NNB	Default	NNB	Default	NNB
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. T. E.	0.012	-2.687	0.022**	-1.872	-0.002	-9.820***	0.007	-1.748	-0.003	-1.945
	(0.03)	(3.89)	(0.01)	(2.28)	(0.01)	(3.46)	(0.01)	(7.01)	(0.00)	(9.11)
Interest rate (Nov.2007)	-0.332	-6.119	-0.033	17.112	-0.083	33.210	-0.009	-12.927	0.053	-102.432
	(0.25)	(40.59)	(0.09)	(29.50)	(0.06)	(36.44)	(0.06)	(81.69)	(0.07)	(119.80)
Income	-0.000	0.000*	-0.000	0.000	-0.000	0.000***	-0.000	0.000	-0.000	-0.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Age	0.001	-0.350***	-0.001***	-0.413***	0.000	-0.563***	-0.000	-0.389**	-0.000	-0.183
	(0.00)	(0.09)	(0.00)	(0.07)	(0.00)	(0.09)	(0.00)	(0.19)	(0.00)	(0.27)
No other cards (%)	0.014	1.699	0.008*	5.471***	-0.001	3.994**	0.001	6.460*	0.002	-6.851
	(0.01)	(2.02)	(0.00)	(1.32)	(0.00)	(1.69)	(0.00)	(3.53)	(0.00)	(4.98)
Baseline Value	0.200***	-24.323***	0.049***	-21.319***	0.035***	-16.866***	0.006*	13.197*	0.009*	49.435***
	(0.02)	(3.74)	(0.01)	(2.20)	(0.01)	(3.36)	(0.00)	(6.81)	(0.01)	(8.71)
Employment Status	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Marital Status	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	4591	4591	12442	12442	17551	17551	7811	7811	4594	4594

Odd columns contain logit models. All estimations are performed on a cross section of clients observed in the month on June 2008. The reported coefficient are marginal effects. Inference is based on a robust variance covariance matrix. An individual is considered a defaulter if either charged-off or after six consecutive months of delinquency. Even columns contain OLS models. All estimations are performed on a panel of clients observed between the months of April and June 2008. Inference is based on a variance covariance matrix clustered by individuals. Net new borrowing is defined as the monthly increase in credit card debt for retail purposes minus the monthly payment and interest rate charges. In all regression, Interest Rate (2007) is the pre-treatment interest rate applied to the client. Income is individual income, measured in pounds. Age is individual age, measured in years. Employment and marital status are a set of binary indicators pertaining to these categories. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. M.D.E. stands for minimum detectable effects. M.D.E. are calculated using sample estimations, and setting $\alpha = 0.05$ and $1 - \beta = 0.80$. Reported, is the value the mean would have to take under the treatment for a significant treatment effect to be detected at the 5% level.

4.5 Conclusions

Using experimental data coming from a unique database, we are able to isolate the causal effect of an increase in interest rate on the probability of defaulting and credit demand for credit card users. Given the experimental nature of the data, we are able to observe a true exogenous variation in the interest rate that does not depend, as it generally happens, on individual characteristics.

Other than identifying a causal relationship between an interest rate increase and default probability, the paper assesses whether this causal effect is heterogeneous across costumers. The answer is that an increase in the interest rate affects individual differently depending on how borrowing constrained they are. Not-constrained individuals face an increase in the interest rate by reducing their credit demand. However, constrained individuals respond to an interest rate increase by worsening their credit position, and eventually defaulting.

We check the robustness of our results by estimating duration models of default probability. As expected, duration models confirm the main results.

Appendix: additional results (chapter 4).

TABLE 4.9: Default Rate in Cells 2 3 5 6.

	(1) Cell 2	(2) Cell 3	(3) Cell 5	(4) Cell 6
Default rate (constant)	0.11^{***}	0.05^{***}	0.04^{***}	0.04^{***}
	(0.01)	(0.00)	(0.00)	(0.00)
Observations	281	137	2317	793

The estimations is performed on a cross section of clients observed in the month of June 2008. The reported coefficient are marginal effects. Inference is based on a robust variance covariance matrix. An individual is considered a defaulter if either charged-off or after six consecutive months of delinquency. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Bibliography

Abowd, J. M. and D. Card (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57, 411–45.

Adams, W., L. Einav, and J. Levin (2009): "Liquidity Constraints and Imperfect Information in Subprime Lending," *American Economic Review*, 99, 49–84.

Addison, J. T., J. Burton, and T. S. Torrance (1984): "Causation, Social Science and Sir John Hicks," *Oxford Economic Papers*, 36, 1–11.

Agarwal, S., S. Chomsisengphet, and C. Liu (2010): "The Importance of Adverse Selection in the Credit Card Market: Evidence from Randomized Trials of Credit Card Solicitations," *Journal of Money, Credit and Banking*, 42, 743–754.

Alan, S. (2006): "Entry Costs and Stock Market Participation over the Life Cycle," *Review of Economic Dynamics*, 9, 588–611.

——— (2012): "Do disaster expectations explain household portfolios?" *Quantitative Economics*, 3, 1–28.

Alan, S. and G. Loranth (2013): "Subprime Consumer Credit Demand: Evidence from a Lender's Pricing Experiment," *Review of Financial Studies*, 26, 2353–2374.

Alessandrini, D. (2014): "On the Cyclicality of Schooling Decisions: Evidence from Canadian Data," Working Paper Series $16_14, The Rimini Centre for Economic Analysis$.

Allison, P. D. (2009): *"Fixed effects regression models"*, SAGE.

Altonji, J. G., A. A. S. Jr., and I. Vidangos (2013): "Modeling Earnings Dynamics," *Econometrica*, 81, 1395–1454.

Andersen, E. B. (1970): "Asymptotic Properties of Conditional Maximum-Likelihood Estimators," *Journal of the Royl Statistical Society.*

Arrondel, L. and H. Calvo-Pardo (2014): "Endogenous non-tradable earnings and households' demand for risky assets," Discussion Paper Series In Economics And Econometrics 1414, Economics Division, School of Social Sciences, University of Southampton.

ARRONDEL, L., H. C. PARDO, AND X. OLIVER (2010): "Temperance in Stock Market Participation: Evidence from France," *Economica*, 77, 314–333.

AUSUBEL (1999): "Adverse Selection in the Credit Card Market," Tech. rep., •.

BAKER, M. (1997): "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings," *Journal of Labor Economics*, 15, 338–75.

BECKER S., O. (2001): "Why don't Italians finish University?" Tech. rep., •.

BERTAUT, C. C. (1998): "Stockholding Behavior Of U.S. Households: Evidence From The 1983-1989 Survey Of Consumer Finances," *The Review of Economics and Statistics*, 80, 263–275.

BETERMIER, S., T. JANSSON, C. PARLOUR, AND J. WALDEN (2012): "Hedging labor income risk," *Journal of Financial Economics*, 105, 622–639.

BETTS, J. R. AND L. L. MCFARLAND (1995): "Safe Port in a Storm: The Impact of Labor Market Conditions on Community College Enrollments," *Journal of Human Resources*, 30, 741–765.

BIS (2013): "Government response to the Bristol University report on high cost credit," Tech. rep., Department for Business, Innovation and Skills.

BONAPARTE, Y., R. COOPER, AND G. ZHU (2012): "Consumption smoothing and portfolio rebalancing: The effects of adjustment costs," *Journal of Monetary Economics*, 59, 751–768.

BONAPARTE, Y., G. M. KORNIOTIS, AND A. KUMAR (2014): "Income hedging and portfolio decisions," *Journal of Financial Economics*, 113, 300–324.

BORELLA, M. (2001): "The Error Structure of Earnings: an Analysis on Italian Longitudinal Data," CeRP Working Papers 07, Center for Research on Pensions and Welfare Policies, Turin (Italy).

BROWNING, M., M. EJRN?S, AND J. ALVAREZ (2010): "Modelling Income Processes with Lots of Heterogeneity," *Review of Economic Studies*, 77, 1353–1381.

BRUNELLO, G. AND L. CAPPELLARI (2008): "The labour market effects of Alma Mater: Evidence from Italy," *Economics of Education Review*, 27, 564–574.

BRUNELLO, G. AND R. MINIACI (1999): "The economic returns to schooling for Italian men. An evaluation based on instrumental variables1," *Labour Economics*, 6, 509–519.

BURGESS, S. AND C. PROPPER (1998): "An Economic Model of Household Income Dynamics, with an Application to Poverty Dynamics among American Women," CEPR Discussion Papers 1830, C.E.P.R. Discussion Papers.

CAMPBELL, J. Y. (2006): "Household Finance," *Journal of Finance*, 61, 1553–1604.

Carneiro, P. and J. J. Heckman (2002): "The Evidence on Credit Constraints in Post–secondary Schooling," *Economic Journal*, 112, 705–734.

Chamberlain, G. (1980): "Analysis of Covariance with Qualitative Data," *Review of Economic Studies*, 47, 225–38.

Chang, Y., J. Hong, and M. Karabarbounis (2013): "Life Cycle Uncertainty and Portfolio Choice Puzzles," Tech. rep.

Cheng, S. and J. S. Long (2007): "Testing for IIA in the Multinomial Logit Model," *Sociological Methods & Research*, 35, 583–600.

Chiappori, P.-A. and B. Salanie (2000): "Testing for Asymmetric Information in Insurance Markets," *Journal of Political Economy*, 108, 56–78.

Christian, M. S. (2007): "Liquidity constraints and the cyclicality of college enrollment in the United States," *Oxford Economic Papers*, 59, 141–169.

Clarke, K. A. (2005): "The Phantom Menace: Omitted Variable Bias in Econometric Research," *Conflict Management and Peace Science*, 22, 341–352.

Cocco, J., F. Gomes, and P. Maenhout (1998): "Consumption and Portfolio Choice over the Life-Cycle," Center for Economic Studies - Discussion papers ces9805, Katholieke Universiteit Leuven, Centrum voor Economische Studien.

Data Monitor UK (2008): "Plastic Cards 2008 Report: Nonstandard Credit Cards in the UK," Tech. rep.

DeJong, D. N. and B. F. Ingram (2001): "The Cyclical Behavior of Skill Acquisition," *Review of Economic Dynamics*, 4, 536–561.

Dellas, H. and P. Sakellaris (2003): "On the cyclicality of schooling: theory and evidence," *Oxford Economic Papers*, 55, 148–172.

Demarinis, G., D. L. Massimo Iaquinta and, and D. Viola (2014): ""Analisi quantitativa della mobilità studentesca negli atenei italiani. Confronto territoriale fra domanda e offerta di formazione universitaria"," Tech. rep., •.

Edelberg, W. (2004): "Testing for Adverse Selection and Moral Hazard in Consumer Loan Markets," Tech. rep., •.

Einav, L., M. Jenkins, and J. Levin (2012): "Contract Pricing in Consumer Credit Markets," *Econometrica*, 80, 1387–1432.

Elmendorf, D. W. and M. S. Kimball (2000): "Taxation of Labor Income and the Demand for Risky Assets," *International Economic Review*, 41, 801–33.

FIASCHI, D. AND C. GABRIELLINI (2013): "Wage Functions and Rates of Return to Education in Italy," Tech. rep.

FINANCIAL CONDUCT AUTHORITY (2014): "Proposals for a price cap on highcost shortterm credit," Tech. rep., •.

GAMBACORTA, L. (2005): "How Do Banks Set Interest Rates?" Temi di discussione (Economic working papers) 542, Bank of Italy, Economic Research and International Relations Area.

GIBBONS, S. AND A. VIGNOLES (2012): "Geography, choice and participation in higher education in England," *Regional Science and Urban Economics*, 42, 98–113.

GINE, X., J. GOLDBERG, AND D. YANG (2012): "Credit Market Consequences of Improved Personal Identification: Field Experimental Evidence from Malawi," *American Economic Review*, 102, 2923–54.

GIORNO, C., P. RICHARDSON, D. ROSEVEARE, AND P. VAN DEN NOORD (1995): "Estimating Potential Output, Output Gaps and Structural Budget Balances," Tech. rep.

GOLLIER, C. (2002): *What does the theory say about household portfolios?, in Household Portfolios*, MIT press, chap. 1.

GOMES, F. AND A. MICHAELIDES (2002): "Life-Cycle Asset Allocation: A Model with Borrowing Constraints, Uninsurable Labour Income Risk and Stock-Market Participation Cost," •.

——— (2005): "Optimal Life-Cycle Asset Allocation: Understanding the Empirical Evidence," *Journal of Finance*, 60, 869–904.

GOURIEROUX, C., A. MONFORT, AND A. TROGNON (1984): "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica*, 52, 701–20.

GROSS, D. B. AND N. S. SOULELES (2002): "An Empirical Analysis of Personal Bankruptcy and Delinquency," *Review of Financial Studies*, 15, 319–347.

GUISO, L. AND T. JAPPELLI (2002): "Private Transfers, Borrowing Constraints and the Timing of Homeownership," *Journal of Money, Credit and Banking*, 34, 315–39.

——— (2005): "Awareness and Stock Market Participation," *Review of Finance*, 9, 537–567.

GUISO, L., T. JAPPELLI, AND D. TERLIZZESE (1996): "Income Risk, Borrowing Constraints, and Portfolio Choice," *American Economic Review*, 86, 158–72.

GUVENEN, F. AND A. SMITH (2010): "Inferring labor income risk from economic choices: an indirect inference approach," Tech. rep.

HALIASSOS, M. AND C. C. BERTAUT (1995): "Why Do So Few Hold Stocks?" *Economic Journal*, 105, 1110–29.

HALIASSOS, M. AND A. MICHAELIDES (2003): "Portfolio Choice and Liquidity Constraints," *International Economic Review*, 44, 143–177.

HAYEK, F. A. (1967): *Studies in Philosophy, Politics and Economic*, Routledge and Kegan Paul.

HEATON, J. AND D. LUCAS (1997): "Market Frictions, Savings Behavior, And Portfolio Choice," *Macroeconomic Dynamics*, 1, 76–101.

HECKMAN, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–61.

HOOVER, K. (2005): "Economic Theory and Causal Inference," Working Papers 64, University of California, Davis, Department of Economics.

ISTAT (2002): ""Universita e lavoro"," .

——— (2014): ""Indagine multiscopo annuale sulle famiglie -Aspetti della vita quotidiana"," .

JALAN, J. AND M. RAVALLION (2001): "Household income dynamics in rural China," Policy Research Working Paper Series 2706, The World Bank.

JAPPELLI, T., M. HALIASSOS, AND L. GUISO (2001): "The profile of European stockholders," *Revue d'Économie Financière*, 64, 163–171.

JENKINS, S. P. (2000): "Modelling household income dynamics," *Journal of Population Economics*, 13, 529–567.

JIANAKOPLOS, N. A. AND A. BERNASEK (1998): "Are Women More Risk Averse?" *Economic Inquiry*, 36, 620–30.

KANE, T. J. (1994): "College Entry by Blacks since 1970: The Role of College Costs, Family Background, and the Returns to Education," *Journal of Political Economy*, 102, 878–911.

KAPETANIOS, G. (2008): "A bootstrap procedure for panel data sets with many cross-sectional units," *Econometrics Journal*, 11, 377–395.

KARLAN, D. AND J. ZINMAN (2009): "Observing Unobservables: Identifying Information Asymmetries With a Consumer Credit Field Experiment," *Econometrica*, 77, 1993–2008.

——— (2010): "Expanding Credit Access: Using Randomized Supply Decisions to Estimate the Impacts," *Review of Financial Studies*, 23, 433–464.

KING, M. A. AND J. I. LEAPE (1998): "Wealth and portfolio composition: Theory and evidence," *Journal of Public Economics*, 69, 155–193.

KNUPFER, S., E. RANATAPUSKA, AND M. SARVIMAKI (2013): "Labour Market Experiences and Portfolio Choice: Evidence from The Finnish Great Depression," Unpublished Paper.

KOCHERLAKOTA, N. R. (1996): "The Equity Premium: It's Still a Puzzle," *Journal of Economic Literature*, 34, 42–71.

KOO, H. K. (1999): "Consumption and portfolio selection with labor income: A discrete-time approach," *Mathematical Methods of Operations Research*, 50, 219–243.

KWAK, D. W. AND J. M. WOOLDRIDGE (2009): "“The robustness of fixed-effects logit estimator to violation of conditional independence assumption," Tech. rep., •.

LANGE, O. (1945): "The Scope and Method of Economics," *The Review of Economic Studies*, 13, 19–32.

LILLARD, L. A. AND Y. WEISS (1979): "Components of Variation in Panel Earnings Data: American Scientists, 1960-70," *Econometrica*, 47, 437–54.

LILLARD, L. A. AND R. J. WILLIS (1978): "Dynamic Aspects of Earning Mobility," *Econometrica*, 46, 985–1012.

LIVSHITS, I., J. MACGEE, AND M. TERTILT (2010): "Accounting for the Rise in Consumer Bankruptcies," *American Economic Journal: Macroeconomics*, 2, 165–93.

LUCIFORA, C., S. COMI, AND G. BRUNELLO (2000): "The Returns to Education in Italy: A New Look at the Evidence," IZA Discussion Papers 130, Institute for the Study of Labor (IZA).

LUSARDI, A. AND P. TUFANO (2009): "Debt Literacy, Financial Experiences, and Overindebtedness," CeRP Working Papers 83, Center for Research on Pensions and Welfare Policies, Turin (Italy).

MACURDY, T. E. (1982): "The use of time series processes to model the error structure of earnings in a longitudinal data analysis," *Journal of Econometrics*, 18, 83–114.

MANACORDA, M. AND E. MORETTI (2006): "Why do Most Italian Youths Live with Their Parents? Intergenerational Transfers and Household Structure," *Journal of the European Economic Association*, 4, 800–829.

MARAVALL, A. AND A. DEL RÍO (2001): "Time Aggregation and the Hodrick-Prescott Filter," Tech. rep.

MASSA, M. AND A. SIMONOV (2006): "Hedging, Familiarity and Portfolio Choice," *Review of Financial Studies*, 19, 633–685.

MATTILA, J. P. (1982): "Determinants of Male School Enrollments: A Time-Series Analysis," *The Review of Economics and Statistics*, 64, 242–51.

MEGHIR, C. AND L. PISTAFERRI (2004): "Income Variance Dynamics and Heterogeneity," *Econometrica*, 72, 1–32.

Meier, S. and C. Sprenger (2009): "Discounting and Defaulting: Evidence from Time Preference Experiments and Administrative Credit Data," Tech. rep.

Merton, R. C. (1971): "Optimum consumption and portfolio rules in a continuous-time model," *Journal of Economic Theory*, 3, 373–413.

Meyer, B. D. (2015): "Household Surveys in Crisis," *Journal of Economic Perspectives.*

Micklewright, J., M. Pearson, and S. Smith (1990): "Unemployment and Early School Leaving," *The Economic Journal*, 100, pp. 163–169.

Miniaci, R. and G. Weber (2002): *Econometric Issues in the Estimation of Household Portfolio Models, in Household Portfolios*, MIT press, chap. 4.

MIUR (2009): ""Gli studenti"," .

Méndez, F. and F. Sepúlveda (2012): "The Cyclicality of Skill Acquisition: Evidence from Panel Data," *American Economic Journal: Macroeconomics*, 4, 128–52.

OFT (2010): "Review on high-cost credit," Tech. rep., Office of Fair Trade.

Ordine, P. and C. Lupi (2009): "Family Income and Students' Mobility," *Giornale degli Economisti*, 68, 1–23.

Pelizzon, L. and G. Weber (2009): "Efficient portfolios when housing needs change over the life cycle," *Journal of Banking & Finance*, 33, 2110–2121.

Perli, R. and P. Sakellaris (1998): "Human capital formation and business cycle persistence," *Journal of Monetary Economics*, 42, 67–92.

Perraudin, W. R. M. and B. E. Sorensen (2000): "The demand for risky assets: Sample selection and household portfolios," *Journal of Econometrics*, 97, 117–144.

Personal Finance Research Centre (2009): "Facing the squeeze," Tech. rep., •.

——— (2011): "Facing the squeeze," Tech. rep., •.

——— (2014): "The impact on business and consumers of a variable cap on the total cost of credit that can be charged in the short to medium term fixed rate high cost credit market," Tech. rep., •.

Petrongolo, B. and M. J. San Segundo (2002): "Staying-on at school at 16: the impact of labor market conditions in Spain," *Economics of Education Review*, 21, 353–365.

PFRC - Bristol (2013): "The impact on business and consumers of a cap of the total cost of credit," Tech. rep., Universiy of Bristol.

Pissarides, C. A. (1981): "Staying-on at School in England and Wales," *Economica*, 48, 345–63.

——— (1982): "From School to University: The Demand for Post-Compulsory Education in Britain," *Economic Journal*, 92, 654–67.

Sakellaris, P. and A. Spilimbergo (2000): "Business cycles and investment in human capital: international evidence on higher education," *Carnegie-Rochester Conference Series on Public Policy*, 52, 221–256.

Sanroman, G. (2015): "Cost and Preference Heterogeneity in Risky Financial Markets," *Journal of Applied Econometrics*, 30, 313–332.

Santos Silva, J., T. Silvana, and W. Frank (2015): "Testing Competing Models for Non-negative Data with Many Zeros," *Journal of Econometric Methods*, 4, 18.

Santos Silva, J. and S. Tenreyro (2006): "The Log of Gravity," *The Review of Economics and Statistics*, 88, 641–658.

Santos Silva, J., S. Tenreyro, and K. Wei (2014): "Estimating the extensive margin of trade," *Journal of International Economics*, 93, 67–75.

Schubert, R. (1999): "Financial Decision-Making: Are Women Really More Risk-Averse?" *American Economic Review*, 89, 381–385.

Staffolani, S. and C. Pigini (2012): "Enrolment Decision and University Choice;of Italian Secondary School Graduates," Working Papers 380, Universita' Politecnica delle Marche (I), Dipartimento di Scienze Economiche e Sociali.

Stiglitz, J. E. and A. Weiss (1981): "Credit Rationing in Markets with Imperfect Information," *American Economic Review*, 71, 393–410.

Terry Long, B. (2004): "How have college decisions changed over time? An application of the conditional logistic choice model," *Journal of Econometrics*, 121, 271–296.

Tumino, A. and M. P. Taylor (2015): "The impact of local labour market conditions on school leaving decisions," ISER Working Paper Series 2015-14, Institute for Social and Economic Research.

Vissing-Jorgensen, A. (2002): "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution," *Journal of Political Economy*, 110, 825–853.

——— (2010): "Consumer Credit: Learning Your Customer's Default Risk from What (S)he Buys," Tech. rep., •.

White, M. J. (2007): "Bankruptcy Reform and Credit Cards," *Journal of Economic Perspectives*, 21, 175–200.

Wooldridge, J. M. (2010): *Econometric Analysis of Cross Section and Panel Data*, vol. 1 of *MIT Press Books*, The MIT Press.

——— (2011): "A simple method for estimating unconditional heterogeneity distributions in correlated random effects models," *Economics Letters*, 113, 12–15.

Yatchew, A. and Z. Griliches (1985): "Specification Error in Probit Models," *The Review of Economics and Statistics*, 67, 134–39.