Dimensions of Well-Being: Earnings, Happiness and Domestic Violence

Thesis

March 8, 2013

Dimensions of Well-Being: Earnings, Happiness and Domestic Violence

Cristina Alexandra Carmo dos Santos

A thesis submitted for the Degree of Doctor of Philosophy to the University College London, the University of London.

To my parents and my loving family.

Declaration

I, Cristina Carmo dos Santos, confirm that:

1. the work presented in this thesis is my own and it has not been presented to any other university or institution for a degree;

2. where information has been derived from other sources, I confirm that this has been indicated in the thesis;

3. chapter one is co-authored with my supervisor Dr. Pedro Carneiro;

4. chapters two and three are sole authored papers.

Abstract

This thesis looks at three important aspects of the well-being of individuals. The first chapter looks at earnings and tries to estimate earnings over the life course accounting for selection. It does so by being silent a priori about the relative productivity of those who stay out of work and instead lets the data speak. Data suggest that non-workers are not always worse than workers, and it also suggests cohort effects are also important when lifecycle profiles do not follow the same people over the whole age range. This chapter also provides an economic model which partly explains how higher productivity individuals may leave the market earlier than low productivity ones. The second chapter looks at another dimension of well-being over the life course. It estimates age-happiness profiles and it focusses more specifically on the identification of linear age effects, in a life satisfaction equation which also includes linear cohort and period effects. As in the first chapter, this chapter also accounts for selective attrition. It finds that cohort effects and selection are important and an adequate account of them changes the age effect on happiness quite substantially. The third chapter looks at domestic violence and tries to find a measure of the cost it has for victims. This is an under-researched area in Economics due to the challenges it presents to the discipline: it questions some of the assumptions often made in the literature about cooperation and efficiency in households; it cannot be easily (if at all) inferred from market behaviour; and data are quite sensitive to gather. We have used a data set designed in the UK, which culminates happiness and income data, and find that costs of violence are often larger than what most households would be able to compensate victims for.

Acknowledgments

In different moments of my PhD, I have been indebted to different people. In one way or another, they all made it happen. I thank Afonso Gonçalves da Silva for gently pushing me into the world of academia. I thank Fundação para a Ciência e Tecnologia for financing the first 4 years of my PhD, under the Sub-Programa Ciência e Tecnologia do Segundo Quadro Comunitário de Apoio, grant number PRAXIS XXI/BD/4920/2001. I thank the 2001 UK recession for not giving me a job. I thank Manuela Angelucci, my aunts Heather and Nina and my late grandmother for their friendship and for being my role models. I thank my parents, my older sister and my whole family for all the support and encouragement, and for forgiving me not being there every time I should have. I thank Paul Anand for gently pushing me into applying for my current lectureship. I thank numerous current and former colleagues, and friends for stimulating conversations and/or invaluable support. These include so many I am sure I am forgetting, but also Miss Piggy and Cocas, Rodrigo, Albert, Ailsa, Evelyn and the dotted tribe, Topi Miettinen, Katrien Stevens, Simone Kohnz, Matthias Parey, Josep Mestres, Julieta Trias, Rita Ginja, Elif Bascavusoglu, Brendon McConnell, Marieke Schnabel, Thibaut Lamadon, Andrew Shepard, Renata Narita and Jelmer Ypma. I thank Lali, Bela and Chico for hiding my glasses, or for complaining, when they felt I should stop working and pay them some attention. I thank Abhishek Chakravarty for being there and for sharing this moment with me. I thank University College London for providing me with the facilities, support and time for me to complete this thesis. I thank John McGlynn for the endless patience with which he dealt with all my questions. I thank the Open University for having given me the time to complete this thesis, and in particular, I thank all of those who not only encouraged me to continue, but also gave me room and support to do so, these include Maureen Mackintosh, Jerome de-Henau, Susan Himmelweit and Roberto Simonetti. I thank Melanie Lührmann, Fernanda Pinto, David Ha, Patrícia Almeida, Suncica Vujic and Natalie Timan for forgiving me all the moments I was too busy to return a phone call and for your friendship. I thank my supervisor Christian Dustmann for seeing it through. And finally, I am particularly indebted to my supervisor Pedro Carneiro for the time, stimulating conversations, and for having believed in me.

Contents

1 Correcting selection in age-earnings profiles **18**

1.1 Motivation 20

1.2 Data 24

1.2.1 The Panel Study of Income Dynamics 24

1.2.2 Earnings, participation and hours over the lifecycle 27

1.3 Empirical strategy and estimation results 31

1.3.1 Estimating wage regression equations 32

1.3.2 Selection, cohort effects and attrition 35

1.4 Results 48

1.4.1 Comparing Age-earnings profiles with the NLSY79 51

1.4.2 Exogenous wage shocks: a discussion 58

1.5 Who works and why? A life cycle model of labour supply 58

1.5.1 The basic lifecycle model with flexible labour supply 61

1.5.2 The lifecycle model with flexible labour supply and deterministic endogenous wages 67

1.5.3 The lifecycle model with flexible labour supply and stochastic exogenous wages 73

1.6 Conclusion 78

2 Revisiting the age-happiness profile: Estimating age, period and cohort effects. **79**

2.1 Introduction 81
2.2 Identifying the effects of age, period and *birth* cohort 85
2.3 Data 89
2.4 Estimation Results 95
2.4.1 Attrition, Unobserved Heterogeneity and Ordinal nature of happiness 102
2.4.2 Using school year cohorts 104
2.5 Conclusion 106

3 Costs of domestic violence: a life satisfaction approach 108
3.1 Introduction 110
3.2 Life Satisfaction Approach 114
3.3 Data 115
3.3.1 Domestic Violence in the UK: a few descriptive results 122
3.4 Estimating a Utility function 129
3.4.1 Specifying the relationship between income and utility 130
3.4.2 Estimation results of life satisfaction equation 132
3.5 Estimating the costs of domestic violence at the individual and at the national level 139
3.6 Conclusion 142

List of Figures

1.1 Age-Earnings Profiles for Male White Heads of Household 20
1.2 Inactivity rates for Male White Heads of Household 21
1.3 Observed age-earnings profiles: PSID and CPS 28
1.4 Inactivity rates: PSID and CPS 29
1.5 How different are workers from non-workers: Fixed effects of White men 38
1.6 How different are workers from non-workers: Fixed effects of White women 40
1.7 Attrition rates for Whites % 42
1.8 How different are attritors from those who stay: Fixed effects of White men 44
1.9 How different are attritors from those who stay: Fixed effects of White women 45
1.10 Average fixed effects by cohort 1919-1966 48
1.11 Age-earnings profiles 1979-1985 49
1.12 Age-earnings profiles 1986-1992 50
1.13 How different are workers from non-workers: using the NLSY 55
1.14 How different are attritors from those who stay: using the NLSY . . . 56
1.15 Age-earnings using NLSY: following cohorts born between 1957 and 1664 57
1.16 Participation rates in the basic model 64

1.17 Hours of work in the basic model 65
1.18 Wage profiles in the basic model 66
1.19 Participation rates in the model with deterministic endogenous wages 70
1.20 Hours of work in the model with deterministic endogenous wages . . . 71
1.21 Wage profiles in the model with deterministic endogenous wages . . . 72
1.22 Participation rates in the model with uncertainty in wages 75
1.23 Hours of work in the model with uncertainty in wages 76
1.24 Wage profiles in the model with uncertainty in wages 77

2.1 The happiness profile in age, following different cohorts 1984-2003 81
2.2 Average number of interviews conducted in each month over the 20-year period 90

1 Weekly Hours of Work: PSID and CPS 145
2 Earnings percentiles by gender and skill group: Whites 146
3 Earnings percentiles by gender and skill group: Blacks 147
4 How different are workers from non-workers: Fixed effects of Black men 149
5 How different are workers from non-workers: Fixed effects of Black women 150
6 Attrition rates for Blacks % 152
7 How different are attritors from those who stay: Fixed effects of Black men 153
8 How different are attritors from those who stay: Fixed effects of Black women 154
9 Average fixed effects by cohort 1919-1966 155
10 Accounting for attrition: Age-earnings profiles 1979-1985 using missing observations 157
11 Accounting for attrition: Age-earnings profiles 1986-1992 using missing observations 158

12 Accounting for attrition and cohort effects: Age-earnings profiles using NLSY and including missing observations 159
13 Consumption profiles in the basic model 163
14 Leisure profiles in the basic model 165
15 Wealth profiles in the basic model 167
16 Consumption profiles in the model with deterministic endogenous wages 170
17 Leisure profiles in the model with deterministic endogenous wages . . 172
18 Wealth profiles in the model with deterministic endogenous wages . . 174
19 Consumption profiles in the model with uncertainty in wages 177
20 Leisure profiles in the model with uncertainty in wages 179
21 Wealth profiles in the model with uncertainty in wages 181

List of Tables

1.1 Characteristics of the eight labour market groups defined by gender, skill and race . . . 26

1.2 Fixed-effects estimation results of wage regression equation . . . 34

1.3 Testing for selective employment and attrition: Whites . . . 41

1.4 Characteristics of the eight labour market groups defined by gender, skill and race: NLSY . . . 52

1.5 Testing for selective employment and attrition using the NLSY . . . 54

1.6 Testing for exogeneity of shocks to wage growth using the NLSY . . . 58

2.1 Average Happiness for all $(age, period)$ combinations - Cohort Table . 92

2.2 OLS estimates of a happiness equation, conditioning on age, birth cohort and calendar time . . . 99

3.1 How vulnerable to current and future domestic violence is the sample? 118

3.2 Percentage of victims of domestic violence in last 12 months . . . 119

3.3 Incidence of domestic violence . . . 122

3.4 Testing for alternative specifications of income in the life satisfaction equation: results from Box Cox specification, parameter λ . . . 131

3.5 Happiness equations: main estimation results . . . 133

3.6 Happiness equations: excluding distress index . . . 137

3.7 Happiness equations: excluding distress index . . . 141

8 Testing for selective employment and attrition: Blacks . . . 151

9 OLS estimates of a happiness equation, conditioning on age, birth cohort and calendar time: first 3 months only 183
10 OLS estimates of a happiness equation, conditioning on age, birth cohort and calendar time: first 4 months only 186
11 OLS estimates of a happiness equation, conditioning on age, birth cohort and calendar time: first 6 months only 189
12 OLS estimates of a happiness equation, conditioning on age, birth cohort and calendar time: balanced panel 192
13 OLS estimates of a happiness equation, conditioning on age, birth cohort and calendar time: individuals who answer in first and last waves 195
14 Accounting for unobserved heterogeneity and the ordinal nature of the dependent variable 198
15 Using schooling cohorts and higher frequency intervals for age and period 199

Introduction

Back in 2003, I presented what would become my transfer seminar paper at a student lunch seminar about variance decompositions of happiness data. This was a paper that allowed the permanent and the transitory components of happiness to be correlated, and identification was achieved by imposing restrictions on the second moments of happiness. I still remember having obtained a persistence of happiness from one period to another of 0.95 and a colleague commented on how that was usually the persistence of consumption data. I remember dismissing that as an irrelevant comment because that was not what the paper was about, but in effect, the duality between consumption or income and happiness has always been a part of my research. In those earlier days, the happiness literature was still infant, but over the past 10 years, it showed several regularities which would not have emerged if these data were mainly driven by individual idiosyncracies and moods. In effect, the implications of this research have grown up to the point of there being suggestions to include happiness in country's national accounts as something to maximise, or of there being countries inviting reports such as the "Report by the Commission on the Measurement of Economic Performance and Social Progress", written by Amartya Sen, Joseph Stiglitz and Jean-Paul Fitoussi (2008). This thesis will hopefully combine what is best about both income and happiness as two characterisations of well-being. An important dimension in the analysis of well-being is also its lifecycle aspect at the individual level. This thesis also explores the challenges of estimating lifecycle earnings and happiness and contributes to the literature by proposing new ways of accounting for selection, attrition and cohort effects. This thesis ends with a reconciliation of income and happiness to calculate the value of domestic violence, an under-researched area in Economics due to the empirical and theoretical challenges it offers to the discipline.

The first chapter of this thesis estimates age-earnings profiles using a fixed effects estimator to account for selection. Earnings profiles and understanding who stays out of the labour market at different moments in life is important for several reasons. First of all, for social inequality reasons. If lower earners stay out of the labour market, the gap in lifetime earnings and wealth they accumulate is larger than if it is the highest earners. What is more, as people get older, and as they have to increasingly provide for themselves in old age, it becomes ever more important to understand these incentives and to estimate the age-earnings profiles accounting for selection into work. The wage regression equation depends on variables which are available whether the individual works or not, or whether he is observed or missing. This simple specification, estimated for groups of individuals who are likely to face different labour markets, allows us to impute a wage to those not working and even to those missing from the panel. In doing so, we remain silent about the relative productivity of workers and non-workers and actually find that selection into employment is not always positive, as is often assumed in the literature. To rationalise these findings, this chapter also simulates a lifecycle model of flexible labour supply under uncertainty which shows how higher productivity individuals (or higher earners) may find it optimal to leave the workforce before lower earners.

The second chapter looks at how happiness evolves over the lifecycle. Self-reported happiness has long been debated as an indicator of well-being and some renowned economists, such as Lord Richard Layard, advocate that it should be part of each country's national accounts. There are several reasons why this measure is problematic, and stem from the fact that it runs counter to the revealed preference mechanism that has dominated Economics for over 100 years. Others criticise it on the grounds that adaptive preferences, or framing effects lead to serious measurement error which invalidates comparisons across individuals and over time. Nonetheless, studies performed in different countries, and in different time periods show several empirical regularities, most of which are quite plausible. One of those empirical regularities is the relation between happiness and age. Most studies have found a U-shaped profile of happiness over the lifecycle and interpreted it as unfulfilled overoptimistic preferences of the young, who later in life see uncertainty in their lives clearing and can live happier lives. However, most of these studies do not account for age effects together with both period and cohort effects, and the age-related coefficients are likely to be biased if all three variables have a direct impact on happiness. Moreover, most of these studies do not account for unobserved heterogeneity nor selective attrition. This chapter reviews the literature on the estimation of age-earnings profiles and on the attempts made to account for all three factors. It also proposes an alternative specification which defines age, year and cohort in yearly intervals and still estimates linear effects of all three variables. This approach relies on longitudinal data and on a particular sampling design where the moment of the interview is exogenous and varies within each period, so that when individuals of a particular birth year are interviewed, they may or may not have had their birthday and completed an additional year of age. The German Socio-Economic Panel was used in this paper, firstly because it has been used extensively to estimate age-happiness profiles, and because it proved to be a data set with adequate sampling design. Accounting for cohort effects using OLS no longer delivers a U-shape profile when we include cohort effects. When we account for individual heterogeneity, using either fixed effects or a fixed effects ordered probit, the age-happiness profile is decreasing.

Chapter 3 estimates the amount of income a victim of domestic violence would forego to be freed of violence by estimating its compensating variation. Utility is measured as self-reported happiness, and is modeled as a function of income and incidence of domestic violence. The dataset used is unique by including both victims and non-victims of domestic violence, together with a rich set of conditioning variables. Most data sets that include information on domestic violence have a very selected sample of victims only that is collected through refugees or through police records. However, the data set I use is cross sectional and income is defined in brackets. Unobserved heterogeneity is accounted for using personality variables and a continuous measure of income is imputed by matching the individuals in the data set with comparable individuals from the BHPS. We also discuss and propose solutions to the endogeneity of both income and domestic violence in the happiness equation. Results suggest that the compensating variation of domestic violence tends to be a very large share of most households' annual income (and a not negligible proportion of national GDP).

Chapter 1

Correcting selection in age-earnings profiles

Abstract

This paper estimates age-earnings profiles using the Michigan Panel Study of Income Dynamics, accounting for selection into employment. A wage regression equation, where log wages depend on age and calendar time, is estimated using fixed effects. This equation is estimated separately for different groups defined according to their gender, schooling and race. The wage of non-workers can be estimated using the predicted log wages for those whose age and a fixed effect are available. Results show that age-earnings profiles when including the potential earnings of non-workers is lower than the observed profile, even though the differences are not significant for all ages. At later ages, we often find workers and non-workers facing similar wages, which casts doubts on the assumption of positive selection into work, so often assumed in similar studies. A fixed effects wage regression equation assumes that unexpected wage shocks are not correlated with individual productivity or the labour supply decisions made in each period, but avoids relying on positive selection into employment. We propose a simple lifecycle model of labour supply with *iid* shocks to wages which explains decreasing participation over age and rationalises the choice of higher earners to leave work earlier than middle income earners.

JEL classification: D01, D91, H31, J22

Keywords: earnings, labour supply, selection equation, fixed-effects.

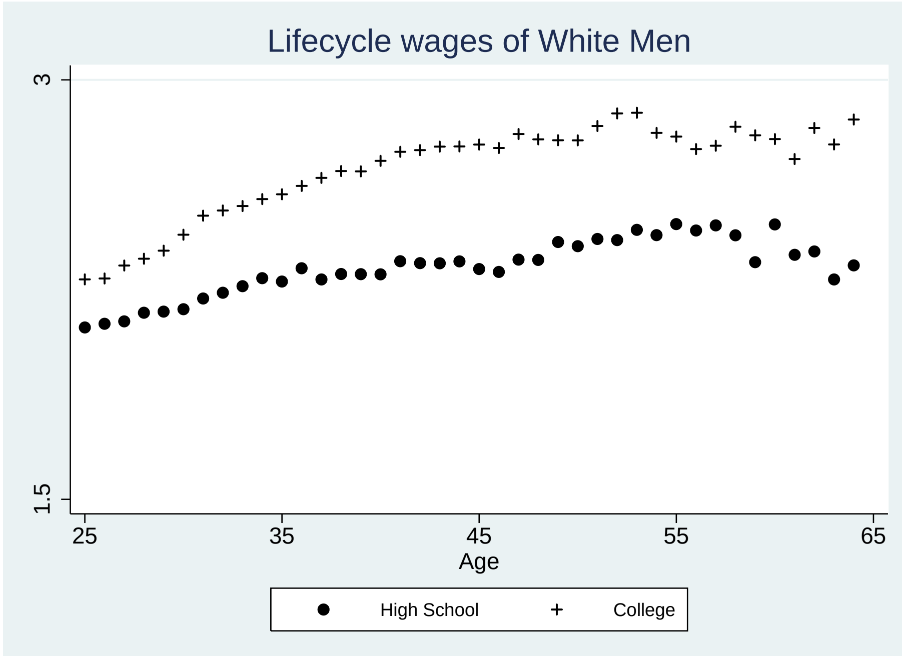


Figure 1.1: Age-Earnings Profiles for Male White Heads of Household

1.1 Motivation

Figure 1.1 shows the *observed* age-earnings profiles for male White heads of household from the Michigan Panel Study of Income Dynamics (PSID) during the 80's and early 90's, according to their schooling. This profile is the result of averaging the wages of individuals observed working at each age. However, and as Figure 1.2 illustrates, the proportion of individuals working at different ages varies substantially. These decreasing participation patterns, specially after age 50, may induce changes in the skill composition of the group observed working, which would lead to a selection into employment bias in these profiles.

We propose a simple way of accounting for selection which relies on a fixed-effects wage regression equation where log wages are defined as a function of age and time. This model assumes that labour supply and potential wages in each period only depend on the individual-specific endowment (which captures initial productivity, preferences and wealth), so that work history and wage shocks do

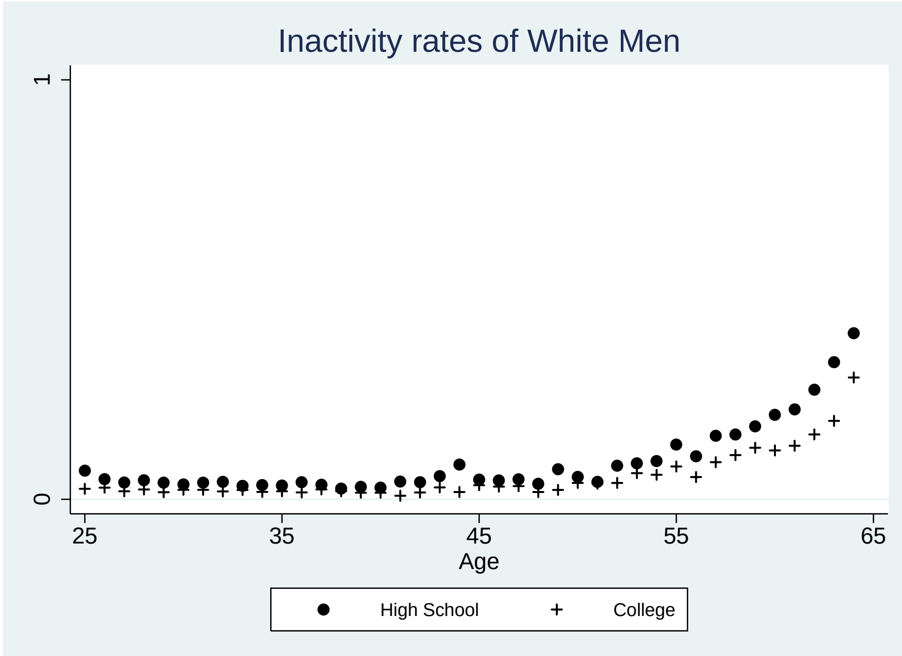


Figure 1.2: Inactivity rates for Male White Heads of Household

not matter[1]. A predicted wage can be imputed to those not working as long as their fixed effect can be estimated, and age can be observed. These imputed wages will be added to observed wages in estimated corrected age-earnings profiles. We find that the difference in productivity between workers and non-workers is not always statistically significant, in particular for older ages when inactivity rates are higher, contradicting the so often used assumption of positive selection into employment(e.g. Petrongolo & Olivetti, 2005; Neal, 2004; Heckman et al., 2000; Chandra, 2003; Blundell et al., 2007).We also account for cohort effects in correcting for age-earnings profiles, and the corrected profiles do not lie below the observed profiles. This is the case particularly for women.

Several other methodologies have been proposed to correct for selection. Most of the literature (e.g. Blundell et al., 2003; Mulligan & Rubinstein, 2004) has hinged on Heckman's structural selection model (1974b) to augment an earnings equation

[1]If we condition our analysis on work experience instead of age, then the strict exogeneity assumption of wage shocks on potential wages is no longer as restrictive and can be implemented. We chose to use age because it is less likely to be measured with error.

with a participation probability term. This procedure is restrictive insofar as it relies on a completely parametric selection and wage determination structure and requires an exclusion restriction which influences labour supply but not wages. The selection equation, even if specified in a semi-parametric way, still relies on a one index type of equation which may be too restrictive if participation decisions depend to some extent on non-wage factors. Fortin (2006) suggests that degrees of greed and work values are strongly associated with labour market outcomes. If these factors are not just a response, but they also contribute to participation decisions and market outcomes, then there is no reason to expect these to increase monotonically with reservation wages. Moreover, the one-index structure cannot model situations where participation decisions are made jointly by couples (e.g. Schafgans & Stelcnery, 2006). On the other hand, in times of growing inequality, such as the 80's in the US (as documented in Juhn et al., 1993), this procedure assumes constant variance. At the other extreme, Blundell et al. (2007) uses bounds to partially identify earnings lifecycle profiles in the UK. However, because employment rates are relatively low in the UK, specially for women, the worst case bounds are not sufficiently informative. This leads the authors to tighten the bounds using several assumptions stemming from Economic theory, namely the assumption that non-workers are lower productivity (and therefore lower earners) than workers. This restriction is often expressed in terms of a median restriction, where it is assumed that the median productivity of workers is larger than the median productivity of non-workers. Petrongolo & Olivetti (2005) uses both matching and imputation methods; relying on the assumption that the ranking of individuals does not vary over time, they impute individual wages by recovering their relative positional wage from the last wave they were observed working; for those that never worked, the authors match their characteristics to predict non-workers' potential wages.

Our model allows us to be silent about the relative or absolute productivity of non-workers vis-a-vis workers. By estimating a wage regression equation for different labour markets, defined according to individual characteristics which do not change over time (we have used gender, highest schooling and race), we can condition our model on important wage and labour supply determinant factors. This wage regression equation allows us to predict a potential wage when the individual is not observed working as long as they have worked at least twice during the period. This is a methodology that works best for individuals with a strong labour market attachment, and cannot correct for selection of those who work very rarely or not at all. This work has however shown that assuming positive selection may not be observed in the data. This methodology can also account for individuals who have left the panel, being one of the few papers which offer a metholodogy which can account for selection and attrition at the same time(exception to be made to Zabel, 1998, who uses a three-equation Heckman's structural selection model); its main restrictions being having wage regression equations specified with covariates that are observed for everybody, even when their wages are not; and having the wage growth rate over the lifecycle independent of unobserved heterogeneity.

Despite its simplicity, this model, and the selection results obtained for different skill, race and gender groups, can be explained by models of intertemporal labour supply with uncertainty and simultaneous labour and consumption decisions which are consistent with our empirical model. The first model of flexible labour supply ever proposed, which assumes labour and consumption are not independent is discussed in Heckman (1974a). It has a closed form solution and shows analytically that in a deterministic lifecycle consumption model where individuals can also freely choose their hours of work, individuals will work more when wages are highest. If wages are anticipated to increase with age (at least up to a certain age as suggested by the *observed* age-earnings profiles in Figure 1.1), then Heckman's model cannot explain the decreasing participation rates observed in Figure 1.2 for any reasonable pure time preference discount rate. Low (2005) shows uncertainty in wages causes individuals to work longer and to consume less at earlier ages in order to build precautionary savings that they can use against future shocks to wages. This buffer stock of wealth will allow individuals to reduce their working hours at later ages, when uncertainty is resolved. We will use a simplified version of Low's model to show specifically how participation decisions depend on individual productivity (measured by a wage level and growth) and on non-labour income. These vary substantially between men and women, and between men and women belonging to different race and skill groups, and shed light on our results. Our model shows that those not working at later ages tend to be both the lowest and the highest earners, for all initial wealth levels. This seems to suggest strong income effects for the highest earners which may also arise in the case of positive matching and reliance on a high-earner spouse (Neal, 2004). We also observe that those with higher non-labour income, are also less likely to work, which explains the differences in participation rates between men and women, but also between White and Black women (as discussed in Neal, 2004, Black women are less likely to be married and have lower spousal income if at all).

The next section describes the data we use and is followed by section 1.3 which describes the empirical model in more detail and presents the estimation results of the fixed-effects wage regression equations. Section 1.4 shows the corrected age-earnings profiles we estimate. Section 1.5 describes the lifecycle models of flexible labor supply we simulate and shows the resulting participation and age-earnings profiles they produce. Section 1.6 concludes.

1.2 Data

1.2.1 The Panel Study of Income Dynamics

Our paper draws on the 1979-1992 period from the Michigan Panel Study of Income Dynamics (PSID). We chose this period and this data set because this period witnessed several changes in participation and income distributions which have been the focus of several studies (e.g. Low, 2005). This panel has been running since 1968 and it became biennial from 1997. Becketti et al. (1988) provides a very good description of the main features and aims of the PSID. This period witnessed important changes in the labor market and wage structures. As documented e.g. in Juhn et al. (1993), the late 80s and early 90s was a period when wage inequality increased substantially, specially when comparing wages received by individuals with different characteristics. It would have been preferable to extend the period to include observations from the late 90's and beyond. However, in 1993, several changes occurred in the way questionnaires are delivered and answers recorded (see e.g. Fitzgerald et al., 1998 for an explanation of these main changes). To avoid including years where additional sources of measurement error could be confounded with cohort effects, we opt to stop our analysis in 1992. We also exclude the Survey of Economic Opportunity (SEO) sample because by only including low income families, it would distort the representativeness of the PSID. In doing so, we have lost more than half of the original individuals.

We estimate wage regression equations for different groups defined according to characteristics which do not change over the time period. These characteristics altogether define the labour market individuals face. We use gender, highest schooling and race. Our analysis requires each group to have enough observations for all ages, but also individuals whose fixed effects can be estimated, i.e. who worked at least twice in the 14 years sampled. To keep enough observations in each group, we use two schooling categories, according to whether individual maximum years of schooling is higher than 12 years of schooling (we call this group the College group) or not, and we only analyse Blacks and Whites. The male sample only includes heads of household, while the female sample includes both heads of household and wives of the heads of household. The age range included is between 25 and 64 years old. This reduced our sample to 7935 individuals, out of which almost 39% responded all years[2]. Table 1.1 shows the number of observations available for each group. The

[2]Some individuals changed their relationship to the head status, or their reported perceived race. Instead of excluding these individuals from the sample, we used the value that was reported at least half of the interviewed periods.

number of observations for the Black groups is relatively low, specially for the College groups of both men and women. Estimating our parameter of interest requires a reasonable number of observations per age group and there are 40 ages within the working age range. For the Black samples, we collapse age into eight 5-year age bands, starting in $[25, 30[$ and going up to $[60, 65[$.

Table 1.1: Characteristics of the eight labour market groups defined by gender, skill and race

	Men				**Women**			
	White		**Black**		**White**		**Black**	
	High School	College	High School	College	High School	College	High School	College
N	1600	1771	210	139	1858	1845	300	212
% wives					84.7	87.8	64.7	78.3
log real hourly wage rates	2.2	2.6	2.1	2.3	1.8	2.2	1.6	2.0
Inactivity rates (%)	8.4	3.8	12.8	8.4	39.1	33.0	41.4	39.1
% women working who are wives					79.8	81.6	53.6	69.5
% individuals w/experience missing	14.5	15.9	18.1	25.9	21.8	22.7	23.7	27.8
Weekly hours of work	46.1	47.2	43.1	44.7	36.0	37.0	38.1	38.8
% individuals lost in estimation	6.7	3.8	10.0	8.6	23.8	16.1	28.0	28.8

We use log real hourly wages as our wage measure. The nominal hourly wages are readily available in the PSID in 1979-1992. These are deflated by the Consumer Price Index. Table 1.1 shows average wages obtained by first computing within-group averages, and then calculating an unweighted average of these. It shows that there are wage differences by gender, skill and race, where men earn more than women, college wages are larger than high school wages, and Whites earn more than Blacks. We define employment status as whether or not a wage is missing in each period. Inactivity rates are calculated by first averaging employment status over the sampled period for each individual and then by computing the unweighted average of these over each group. We tried to explore the reliability of the labour force status employment variable, but this variable has several inconsistent reports over time and as Table 1.1 shows, it is not available for a large proportion of any group sample. Women's inactivity rates are much higher than men's, and not surprisingly, women are less likely to work when they are wives than when they are heads of household (this can be seen by comparing the percentage of women who are wives with the percentage of those who are wives only amongst the working group; in the second group, the incidence of wives is lower). The table also shows that College groups are more likely to work than High School groups, and Whites more likely to work than Black groups. These differences across groups are confirmed by the average weekly hours of work each group has, except for Black Women, who work more hours than their White counterparts, once in employment. These very high inactivity rates still allow us to impute wages for those periods where individuals are not working for a large fraction of the individuals belonging to each group. For men, this percentage is at least 90% for all groups. For women however, and even including wives of the heads of household, this percentage is sometimes as low as 70%.

1.2.2 Earnings, participation and hours over the lifecycle

Figure 1.3 shows the observed average age-earnings profiles calculated for all eight groups using log real hourly wage rates as the measure of earnings. It shows that the age-earnings profiles for all groups are hump-shaped, which can be a result of compositional effects as the proportion of individuals who stay in employment decreases sharply for all groups, as Figure 1.4 shows. College White Men are not only the highest earners, but also the earners with the highest wage growth rate over the lifecycle.

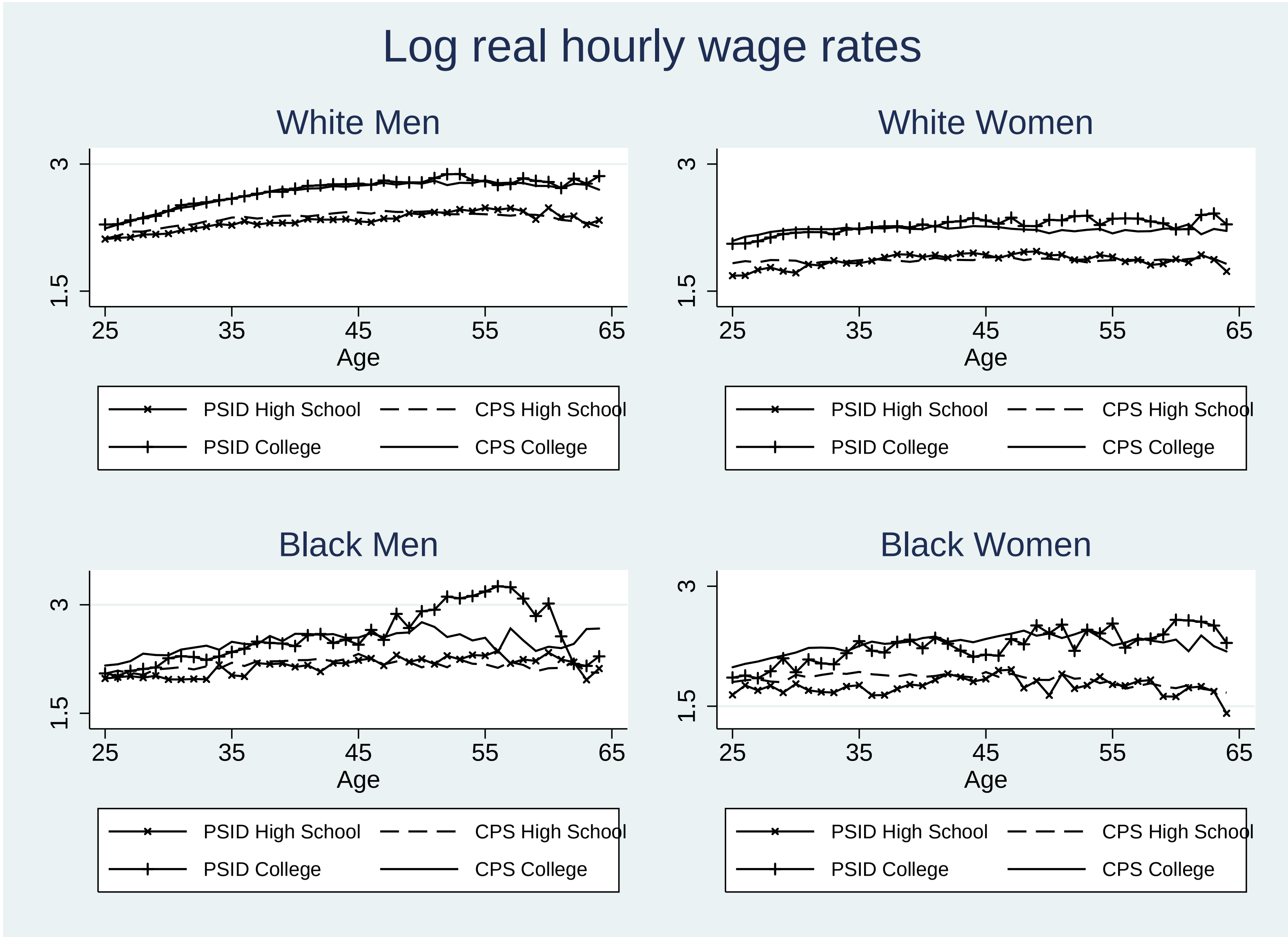


Figure 1.3: Observed age-earnings profiles: PSID and CPS

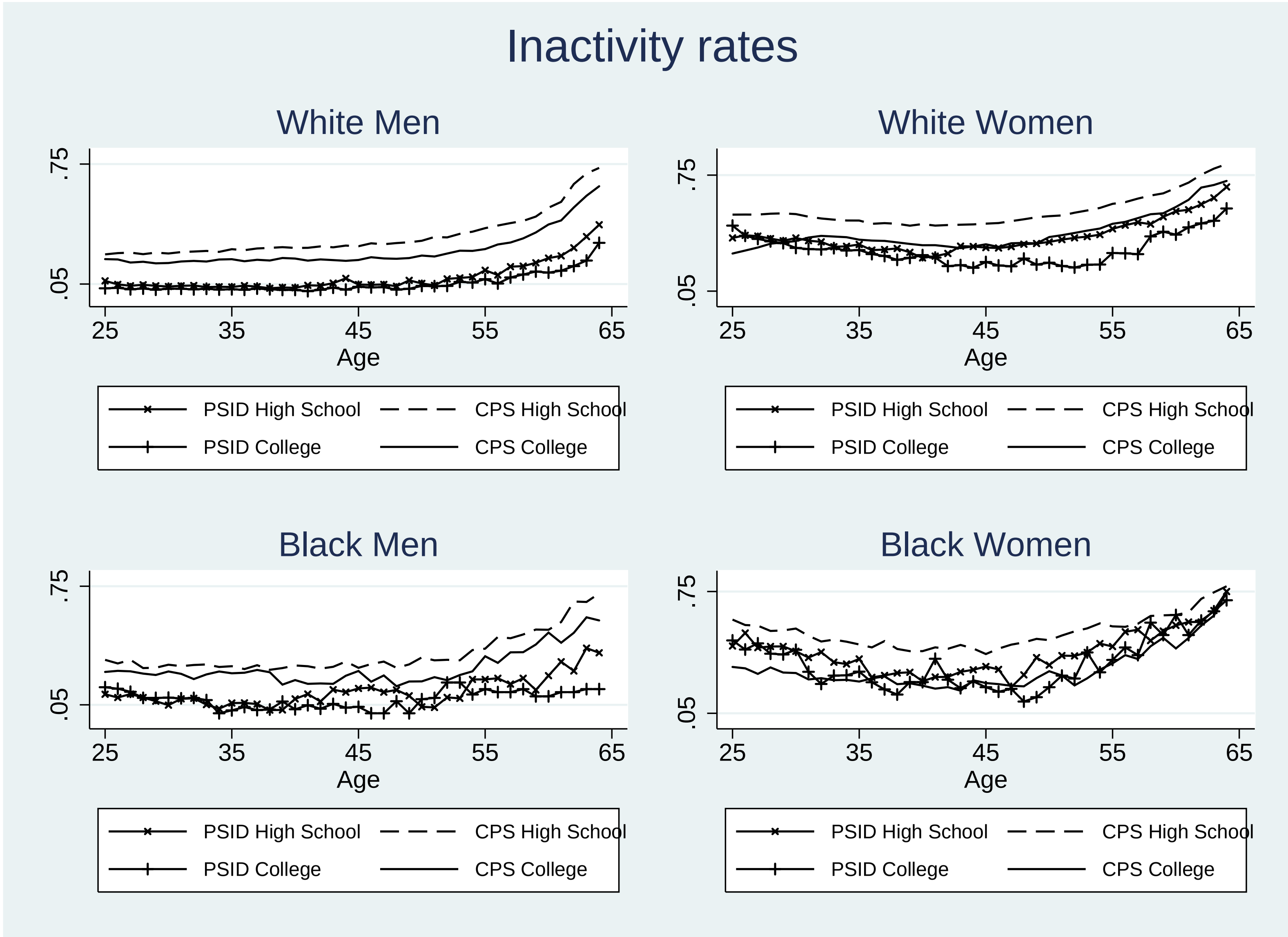


Figure 1.4: Inactivity rates: PSID and CPS

We compare our wage and inactivity measures with the ones resulting from Census data. We use the March Current Population Survey (CPS) waves from a comparable period and sample. We only include the sample period from 1979 until 1992, White and Black men and women aged between 25 and 64, and we only use male heads of household while using both heads and wives of heads for women. CPS does not include a measure of hourly wage rates, but these were derived from annual earnings and a measure of total annual hours (the product of weeks of work and the number of hours worked per week). Our employment variable is still a function of whether the wage is missing or not.

Observed *average* wages are similar in the two data sets, but there is a much larger proportion of individuals in the CPS which are not observed working, and the difference in inactivity rates according to schooling is much higher in the CPS than in the PSID. The apparent similarity between the two data sets in terms of average wages overlooks some interesting differences across the wage distribution. Appendix A shows how different percentiles of the log hourly wage rate compare across the two data sets. As found in previous studies, the magnitudes of observed wages are very different between the PSID and the CPS, but the trends are very similar (See e.g. Gottschalk & Moffitt, 1992; Handcock et al., 2000). However, it is also often the case that the first 20% of the individuals are not earning in the CPS or earning considerably less than in the PSID, but the PSID has a longer upper tail. Appendix A also shows that conditional on working, CPS workers also work considerably less hours than in the PSID. Handcock et al. (2000) point out to differences in the questionnaires and questionnaire delivery as possible factors underlying these discrepancies. Given the dimension of the CPS as a representative cross sectional data set of the US population, it makes more use of proxy and phone interviews than longitudinal surveys. It does not ask as many questions about work and earnings as the PSID, which may contribute to measurement error and, for individuals with lower labour market attachment, significant underreporting of weeks worked or earnings. The notion of headship is also different between data sets, with the PSID having a more conventional stronger notion of head of household[3].

1.3 Empirical strategy and estimation results

We want to estimate the age-earnings profile $E(\log w\,|a), \forall a$, faced by an individual falling into one of the groups defined by gender, race and schooling. However, we only observe the earnings of those who work, i.e for whom employment status e takes the value 1 $E(\log w\,|a\,, e=1)$. Given the relation between these two parameters in Eq. 1.1

$$\begin{aligned} E(\log w|\,a) &= \quad E(\log w|\,a, e=1)\,P(e=1|\,a) \\ &+ \quad E(\log w|\,a, e=0)\,(1-P(e=1|\,a)), \end{aligned} \tag{1.1}$$

observed age-earnings profiles will only be similar to the true age-earnings profiles when there is no selection into employment and both workers and non-workers have the same expected wages, when $E(\log w\,|a\,, e=1) = E(\log\ w\,|a\,, e=0)$. Alternatively, if the proportion of non-workers $(1-P(e=1|\,a))$ is negligible, the weight of the earnings of non-workers in the corrected profile will be so low that significant differences between workers and non-workers will not have a visible impact on the corrected age-earnings profiles. While this proportion is low at early ages, Figure 1.4 has shown us that this is not the case for any group when they get older.

Our wage regression equation assumes log real hourly wages are a function of age a, time t and a fixed effect f and is specified in Eq. 1.2. We include age and calendar time dummy variables, where time is divided into two sets of seven years,

[3]The concept of headship in the CPS is different from the PSID. The CPS discontinued the notion of a head of household in 1980 because of the more equal sharing and social changes taking place in the US. As such, the householder is the person responsible for paying the rent and, in case a couple is responsible for this, then either can be classified as the householder. The PSID kept the original definition of a head of household for consistency sake, which in married couples, means that only under serious bereavement of the male, will he not be the head of household. Given that the PSID is a longitudinal survey whereas the CPS does not allow us to track the same individuals through time, this also means that there will be a higher share of male breadwinner households in the PSID because these were more common when the PSID was first launched. The effect this difference would have on earnings and work patterns is however unclear.

1979-1985 and 1986-1992. The fixed effect provides a measure of productivity and also captures important determinants of wages which can be assumed fixed. These include initial wealth, unobserved ability and preferences for work.

$$\log w_{it} = \alpha_0 + \alpha_a 1\left(a_{it} = a\right) + \alpha_t 1\left(t \geq 1986\right) + f_i + u_{it} \tag{1.2}$$

Averaging the individual predicted log wages for each particular age, after conditioning on time, whether the individual works or not, recovers our parameter of interest $E\left(\log w \left| a\right.\right)$. The crucial advantage of this specification is that we can observe the age and the calendar time for all individuals and for all years sampled, and not just for the workers. In fact, we can even observe these variables for those individuals who have left the panel. However, we are assuming that α_a does not depend on f, i.e, that all individuals within the same (gender, race, skill) group, face the same wage growth regardless of their unobserved initial productivity. All shocks to employment and to productivity are therefore independent of initial conditions. Section 1.4.2 will discuss the plausability of this assumption.

1.3.1 Estimating wage regression equations

Table 1.2 shows the results of our fixed-effects wage regressions, according to Eq. 1.2, for all the eight groups defined by gender, skill and race. For presentational purposes, we only present a subset of the coefficients for Whites. For Blacks, the coefficients of all age intervals are presented.The age coefficients confirm what the observed wage profile figures already hinted at. College groups face age earnings profiles which start higher and are steeper, even if disaccelerating at later ages, than High School groups'. Profiles tend to have an inverted U-shape except for Black women, whose profiles for both skill groups are increasing (even if starting at lowest level than all other groups within the same skill group). For High School men instead, both Black and White, profiles seem mostly flat over the lifecycle. Mid-aged High School White women earn more than at other ages. Year effects only have a significant effect for

White groups, and show wages increase in the second half of the sample period. For Blacks, earnings seem to have stagnated.

Table 1.2: Fixed-effects estimation results of wage regression equation

	White				Black			
	Men		**Women**		**Men**		**Women**	
	High School	College	High School	College	High School	College	High School	College
age 30	0.036	0.137***	-0.021	0.146***	-0.027	0.123*	0.102*	0.123**
	(0.027)	(0.025)	(0.037)	(0.033)	(0.041)	(0.050)	(0.046)	(0.046)
age 36	0.078*	0.276***	0.084	0.207***	0.04	0.217**	0.09	0.308***
	(0.033)	(0.029)	(0.043)	(0.039)	(0.057)	(0.070)	(0.066)	(0.066)
age 41	0.061	0.376***	0.157**	0.309***	-0.093	0.261**	0.204*	0.370***
	(0.040)	(0.035)	(0.051)	(0.048)	(0.080)	(0.094)	(0.084)	(0.091)
age 46	0.015	0.420***	0.219***	0.421***	-0.174	0.327**	0.331**	0.521***
	(0.049)	(0.042)	(0.062)	(0.060)	(0.104)	(0.121)	(0.110)	(0.121)
age 51	-0.005	0.416***	0.211**	0.527***	-0.122	0.594***	0.500***	0.564***
	(0.057)	(0.052)	(0.072)	(0.073)	(0.125)	(0.161)	(0.130)	(0.155)
age 56	-0.026	0.349***	0.252**	0.573***	-0.163	0.548**	0.502***	0.634***
	(0.064)	(0.061)	(0.082)	(0.086)	(0.146)	(0.201)	(0.152)	(0.186)
age 61	-0.062	0.293***	0.161	0.386***	-0.211	0.418	0.566**	0.805**
	(0.073)	(0.070)	(0.093)	(0.098)	(0.167)	(0.232)	(0.176)	(0.244)
late80s	0.031*	0.054***	0.081***	0.046**	0.056	-0.015	-0.058	-0.035
	(0.014)	(0.013)	(0.018)	(0.018)	(0.037)	(0.046)	(0.040)	(0.044)
Constant	2.256***	2.311***	1.678***	1.955***	2.134***	2.201***	1.578***	1.921***
	(0.028)	(0.024)	(0.038)	(0.030)	(0.045)	(0.046)	(0.052)	(0.040)
R^2	0.007	0.05	0.022	0.036	0.01	0.026	0.016	0.039
No. individuals	1493	1703	1416	1548	189	127	216	151

For Whites, age coefficient estimates refer to a single age; for Blacks, it refers to the age group which starts with that value.
For Whites, 25 year old dummy variable omitted; for Blacks, the age group between 25 and 29 years old omitted.
Significance levels: * 5% ** 1% *** 0.1% Standard errors in parentheses

These profiles differ from the observed age-earnings profiles we saw in Figure 1.3, which were all mostly concave. Our accounting for fixed effects will have conditioned the age-earnings profiles on time invariant determinants of labour supply and labour market outcomes. At the same time, fixed effects also accounted for cohort effects, and our empirical estimation of age-earnings profiles will attempt to disentangle cohort from selection effects. These results already show a large difference across groups in terms of initial wage level and wage growth over the lifecycle, two important factors explaining participation decisions, as we will see in Section 1.5.

1.3.2 Selection, cohort effects and attrition

The previous estimation results suggest that estimated age-earnings profiles not only vary substantially across labour market groups, but also differ substantially from their observed counterparts. These differences are due to our account of fixed effects in our wage model. These age coefficients have been estimated without the influence of time invariant individual attributes which condition labour market outcomes, while their influence affects the profiles in Figure 1.3. Analysing the fixed effects of workers and non-workers will allow us to make statements about the relative productivity of these two groups. However, there are two confounding factors in this estimation. These fixed effects are estimated with an unbalanced panel, and therefore an overrepresentation of individuals who stayed in the panel for longer. Secondly, each age-specific average fixed effect is estimated using individuals from a different set of cohorts. This section will compare the fixed effects of workers and non-workers and it will discuss the relative importance attrition and cohort effects seem to have with these data.

Comparing workers and non-workers

To understand whether selection plays a significant role in shaping the estimated age-earnings profiles, Figures 1.5 and 1.6 show age-specific fixed effects averaged for

White male and female respectively. They show the fixed effects of the working, the non-working and the total observed sample. These profiles were computed for each of the sample periods separately, but there do not seem to be important differences in the average fixed effects over the lifecycle according to the macroeconomic context in which individuals entered the labour market. As seen in Eq. 1.1, the expected value of the fixed effects for the whole sample $E(\log w\,|a)$ will be similar to the expected value of the observed sample $E(\log w\,|a\,, e=1)$ when there is little difference between workers and non-workers in their productivity, or when the proportion of non-workers is so small that the impact they have in the overall profile is negligible. Figure 1.5 shows that for White men, the average fixed effects of the observed and of the total interviewed sample follow a very similar pattern, and differences between the two profiles only seem to be significant in later ages for the High School group in the first half of the period.

The sample of non-workers does seem to fare worse than workers throughout the life course, except at the end of the working life, when inactivity rates are higher and the selection term in Eq. 1.1 more important. For women, and because the proportion of women staying out of the labour force at any one period is larger, Figure 1.6 shows a slightly larger difference between the average fixed effects of the observed and the total sample. The differences between workers and non-workers throughout the lifecycle seem to be the same as for White men: while there seems to be positive selection into employment for mid-ages, this is not the case for later ages. χ^2 tests comparing the average fixed effects of workers with non-workers, for the first and the second half of the period, do show that this difference is significant for all four White groups (first two columns of Table 1.3).

Due to a smaller sample size of non-workers for all ages, average fixed effects of non-workers vary more across the lifecycle and confidence bands are also larger[4]. But broadly, Figures 1.5 and 1.6 suggest that selection into employment later in life

[4]Confidence bands calculated, for each age a as $\bar{f}_a \pm 1.96\left(\frac{1}{N_a}\sqrt{\sum_i f_i}\right)$

is not positive. The reasons why some individuals retire earlier than others does not seem to be driven by their productivity.

For the Black groups, conclusions are very similar and results presented in Appendix B.

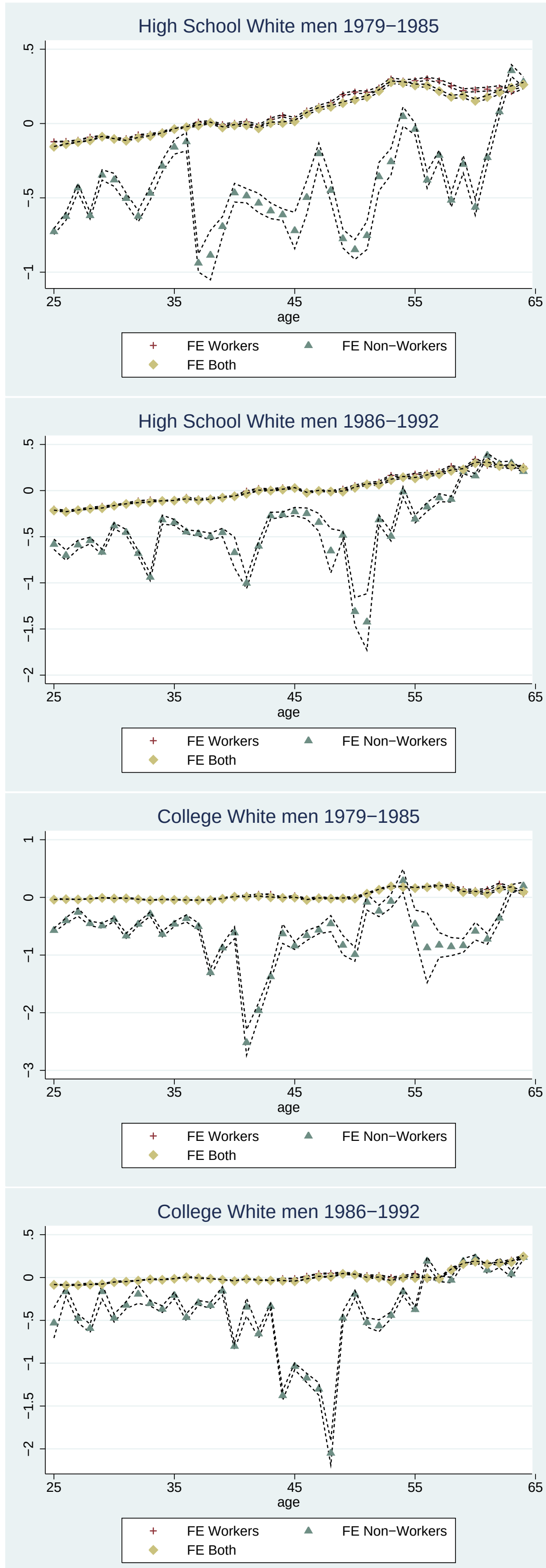


Figure 1.5: How different are workers from non-workers: Fixed effects of White men

The role of attrition and cohort effects

Figures 1.5 and 1.6 showed how selection can be a significant phenomenon and still differences between the fixed effects of workers and of the total sample are small. However, they also showed that these fixed effects change over the lifecycle. For men, these increase with age for the High School group, or they remain relatively constant over time and close to zero for the College group. For women, fixed effects of these two groups decrease with age. For all groups except the group of College men, we have a profile of average fixed effects of the total sample which could not have been produced by a balanced panel with 40 periods[5]. These changes in the average fixed effects occur either because of selective attrition, or because of cohort effects.

If we want to analyse attrition, we have to extend the panel of interviewed people and impute a fixed effect to the years when individuals already left the panel. In this context, where we will have for each age the observed sample of stayers interviewed and the missing sample of attritors, our parameter of interest is defined in Eq. 1.3.

$$\begin{aligned} E\left(\log w \mid a\right) &= E\left(\log w \mid a, e=1, \text{stayer}\right) P\left(e=1, \text{stayer} \mid a\right) \\ &+ E\left(\log w \mid a, e=0, \text{stayer}\right)\left(1-P\left(e=1, \text{stayer} \mid a\right)\right) \\ &+ E\left(\log w \mid a, e=1, \text{attritor}\right) P\left(e=1, \text{attritor} \mid a\right) \\ &+ E\left(\log w \mid a, e=0, \text{attritor}\right)\left(1-P\left(e=1, \text{attritor} \mid a\right)\right) \end{aligned} \tag{1.3}$$

Figure 1.7 shows the attrition rate of the White groups in both sample periods. While negligible in earlier years of the panel, attrition rates are as high as 20% for some age groups in the 1986-1992 period. Attrition rates are larger for Blacks, as can be seen in Appendix B.

[5] By construction, the age-specific average fixed effect profile for the whole sample should be a flat line around zero.

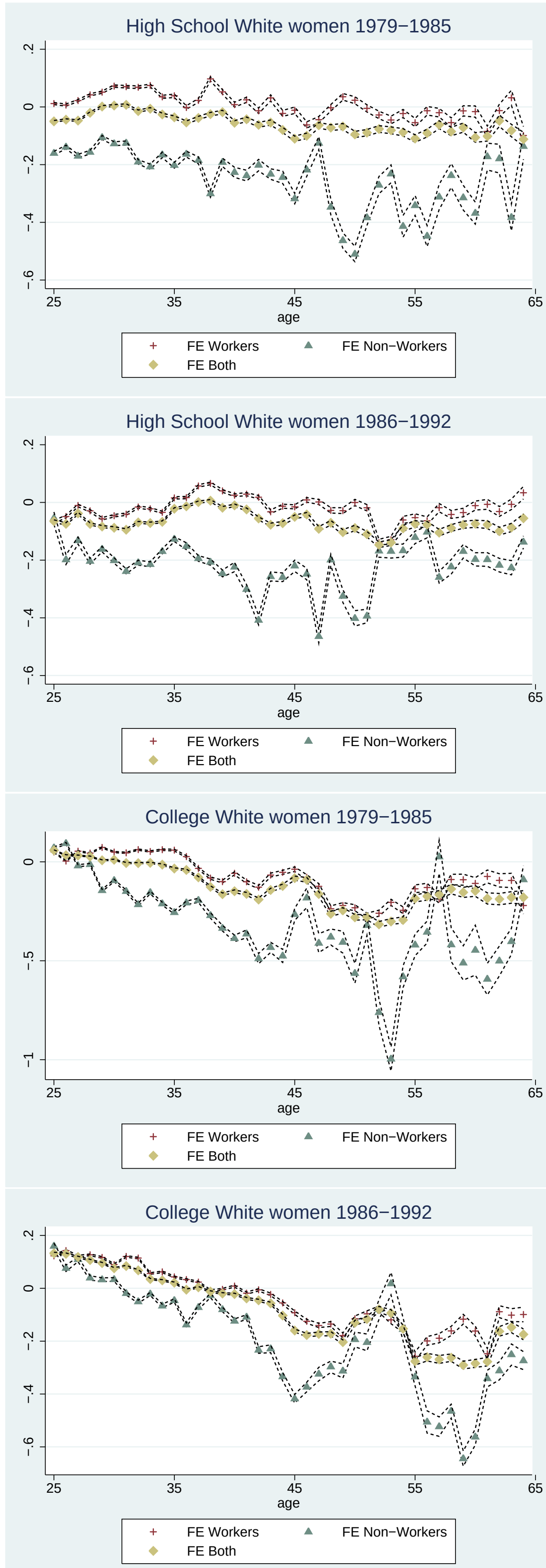


Figure 1.6: How different are workers from non-workers: Fixed effects of White women

Table 1.3: Testing for selective employment and attrition: Whites

	$H_0 : \overline{FE}_w - \overline{FE}_{Nw} = 0$		$\overline{FE}_{\text{obs}} - \overline{FE}_{\text{atr}} = 0$		$\overline{FE}_w - \overline{FE}_{\text{exc.w}} = 0$	
	χ^2	***p*-value**	χ^2	***p*-value**	χ^2	***p*-value**
High School White men						
1979-1985	1054.5	0	301.7	0	932.7	0
1986-1992	1367.6	0	444.5	0	789.6	0
College White men						
1979-1985	934.5	0	227.8	0	789.8	0
1986-1992	1453.7	0	422.4	0	587.1	0
High School White women						
1979-1985	317.6	0	116.1	0	310.7	0
1986-1992	387.2	0	215.5	0	306.0	0
College White women						
1979-1985	481.0	0	110.3	0	409.5	0
1986-1992	404.6	0	231.8	0	363.9	0

Results based on the χ^2 test $(R\varphi)\left(R\text{CovR}'\right)^{-1}(R\varphi)' \sim \chi^2_q$, where φ is the vector of estimated fixed effects R is the projection matrix that transforms the fixed effects into a difference in means between two groups Cov is the covariance matrix of the fixed effects and q is the number of restrictions, one for each age level.

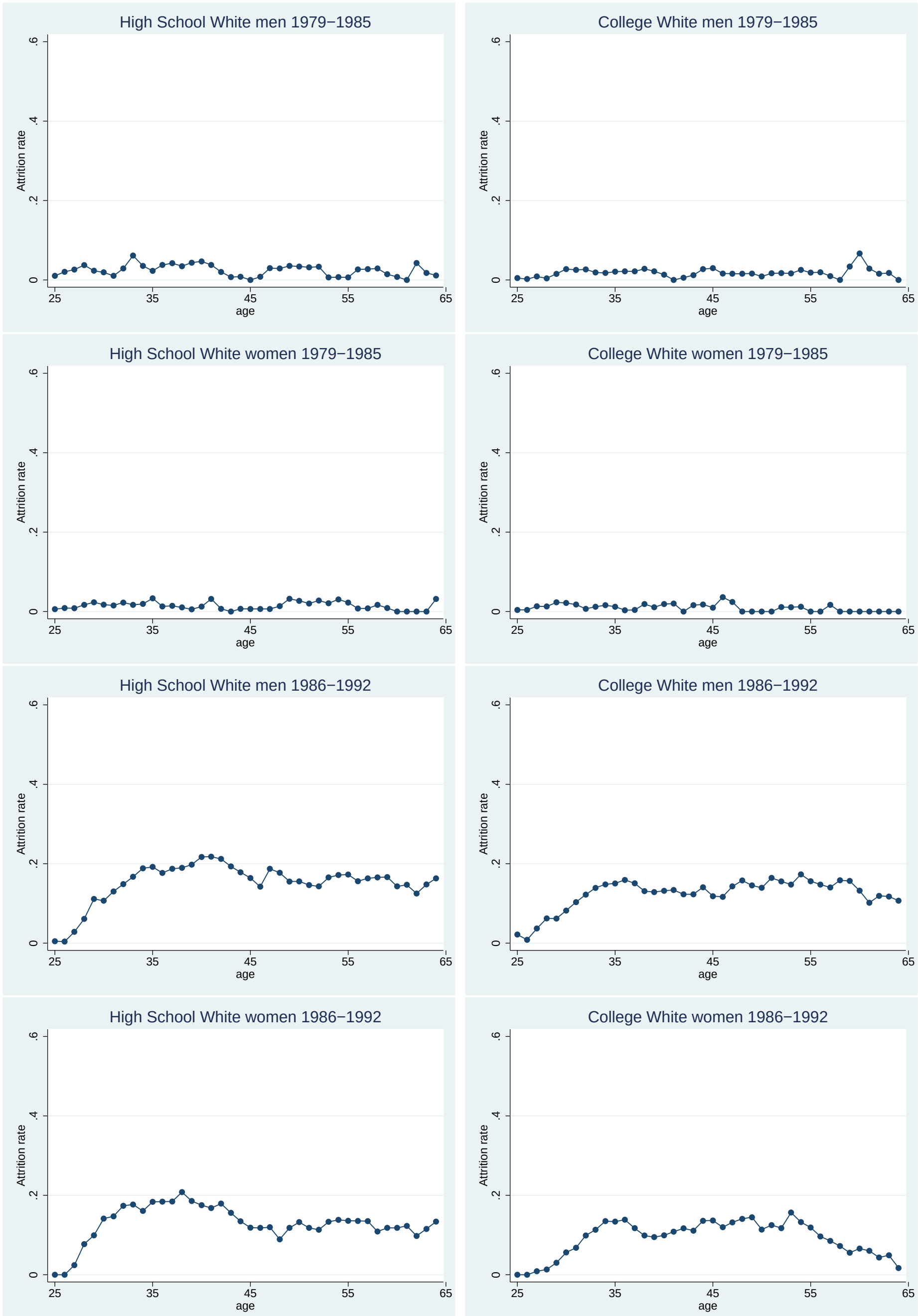


Figure 1.7: Attrition rates for Whites %

Differences in the FE between stayers and attritors would be reflected in a difference in the profile which uses the whole set of observations (both stayers and attritors) and the profiles of those who stay or those who attrite. We have seen that conditioning on employment status does not affect the expected value of the fixed effect $E\left(f\,|a\right)$ (mainly because the selection term is too small when workers are very different from non-workers). A comparison of the average fixed effects between observed workers (instead of all observed), attritors and the whole set of observations (including attritors) will inform us on the relative importance of selective attrition. Figures 1.8 and 1.9 show us the results for White men and women.

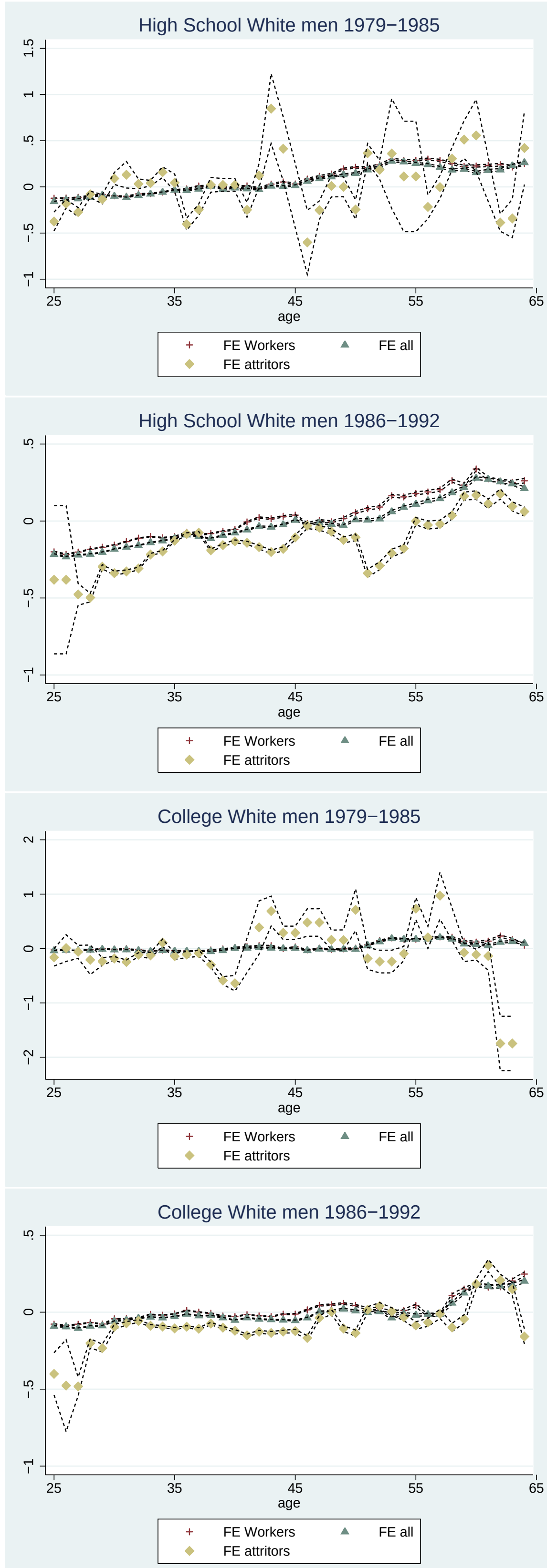


Figure 1.8: How different are attritors from those who stay: Fixed effects of White men

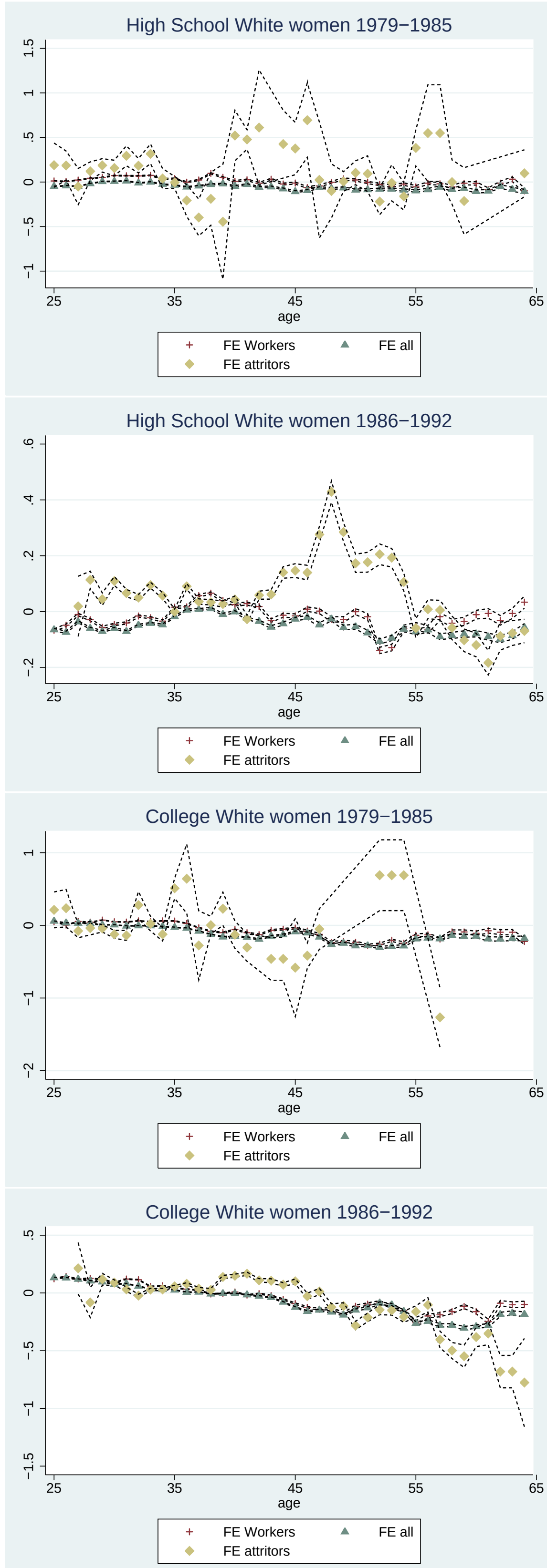


Figure 1.9: How different are attritors from those who stay: Fixed effects of White women

Average fixed effects of observed workers (similar to the profile of all observed individuals) match closely the profile which uses the whole set of observations, including the attritors. This shows selective attrition does not have a substantial impact in the estimation of lifecycle fixed effects and wages, even though attritors are significantly different from workers or from observed individuals (as shown in Table 1.3, second test statistic and p-value). Table 1.3 also shows the χ^2 and the p-value of a test which compares individuals observed working with the remaining groups, attritors included. Results suggest no significant differences between these groups. A joint account of attrition and selection does not seem to have a significant impact on estimated fixed-effects profiles. However, when we look at these three groups across age, we can see significant differences between workers and attritors. For men in the second half of the period, attritors seem to be worse than workers for most ages, while this is not the case in earlier years of our sample. For women, the pattern is reversed and attritors are significantly more productive than workers in the second half of the period. So in effect, for women, the positive effect of attrition on average fixed effects of the total sample cancelled out the negative effect of non-workers, so that the profiles of workers and the total sample are now closer than they were when missing attritors had not been accounted for (Figure 1.6). Most importantly, these Figures and Table 1.3 show that the profile which accounts for attrition does not seem to be any closer to the flat line around zero for most groups. This suggests that attrition is not one of the major factors underlying the variability of estimated age-specific average fixed-effects.

The results for Blacks are in Appendix B. For College groups in the first half, attritors are not significantly different from stayers, but this is mainly due to the small sample of College Black people in earlier years. Attritors tend to be worse than stayers for all groups and periods.

The remaining influence on this estimation, which can be confounded with selection effects are cohort effects. Figure 1.10 shows how average fixed effects change across cohorts. It shows that average fixed effects change significantly with cohort and this relation also varies with the group (Figure 9 in the appendix shows the average fixed effects by cohort for the Black groups). There are stark differences between men and women; while the fixed effects of later born women are larger than those of earlier born women, except for the group of High School White women, for men that is reverse, except for College Black men.

Figures 1.10 and 9 do suggest that cohort effects are the main reason why age-specific average fixed effects do not average to zero for all ages, for both men and women. Regarding men, because later born men have on average lower fixed effects, that pulls the average fixed effect of younger ages in Figure 1.8 down (and of older ages up). This explains the estimated increasing fixed effects of the total sample. The opposite occurs for women.

Our approach does not allow us to separately identify cohort and selection effects. The next section will discuss the selection-corrected profiles we obtain by including the potential wages of non-workers in the corrected profile. We will also estimate these corrected profiles using the NLSY which follows a set of cohorts through their adult life in an attempt to attenuate the impact of cohort effects in our selection-corrected profiles. In order to account for attrition, we will use predicted wages for every observation, including those observations of individuals who dropped the panel. These results are however in Appendix C.

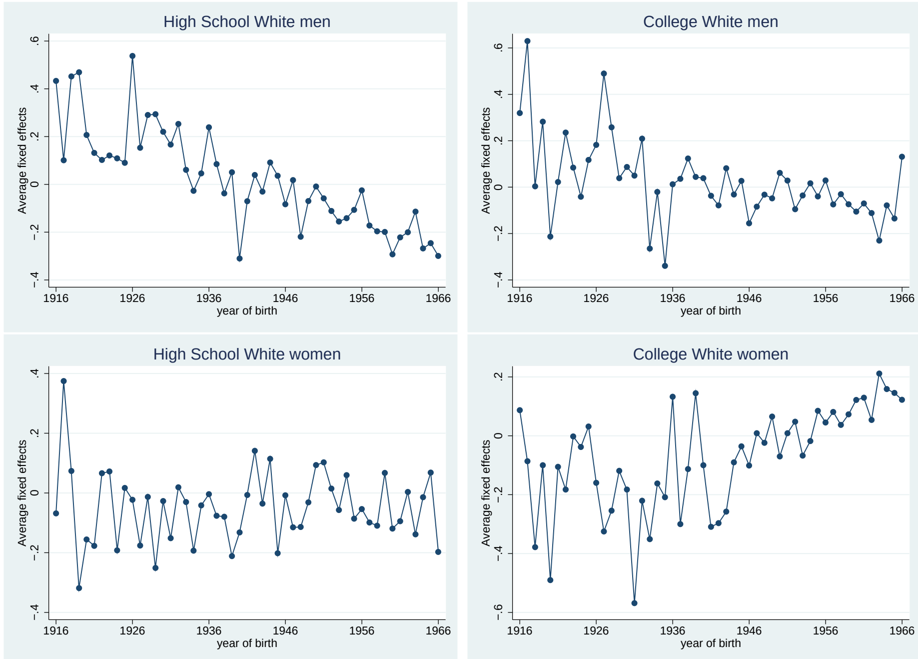


Figure 1.10: Average fixed effects by cohort 1919-1966

1.4 Results

This section shows the impact differences in fixed effects across workers and non-workers have on estimated age-earnings profiles. The corrected profiles in this section follow Eq. 1.1. The corrected profiles include the potential wages of non-workers, estimated using a regression equation which accounts for fixed effects (and therefore cohort effects). These profiles will thus show both how non-workers compare to workers (selection effect) over the lifecycle, but this effect will be confounded by how earlier-born cohorts compare to later-born cohorts (cohort effects). We have seen in Figure 1.10 that the fixed effects of later born women tend to be higher than the fixed effects of earlier born women, and this is the reverse for men. Figures 1.11 and 1.12 show the results for all eight groups in each sample period respectively. Profiles which further account for attrition are in Appendix C.

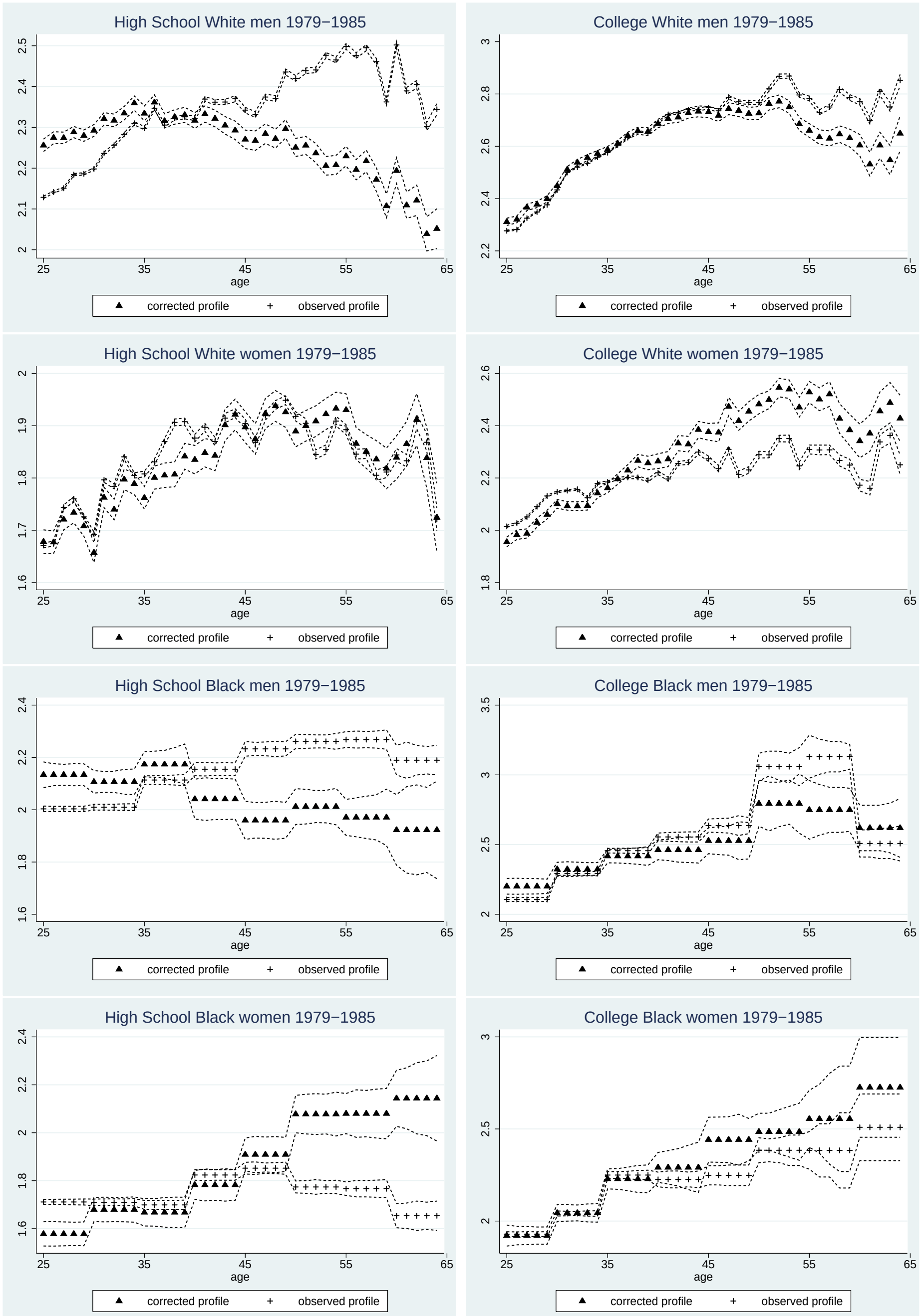


Figure 1.11: Age-earnings profiles 1979-1985

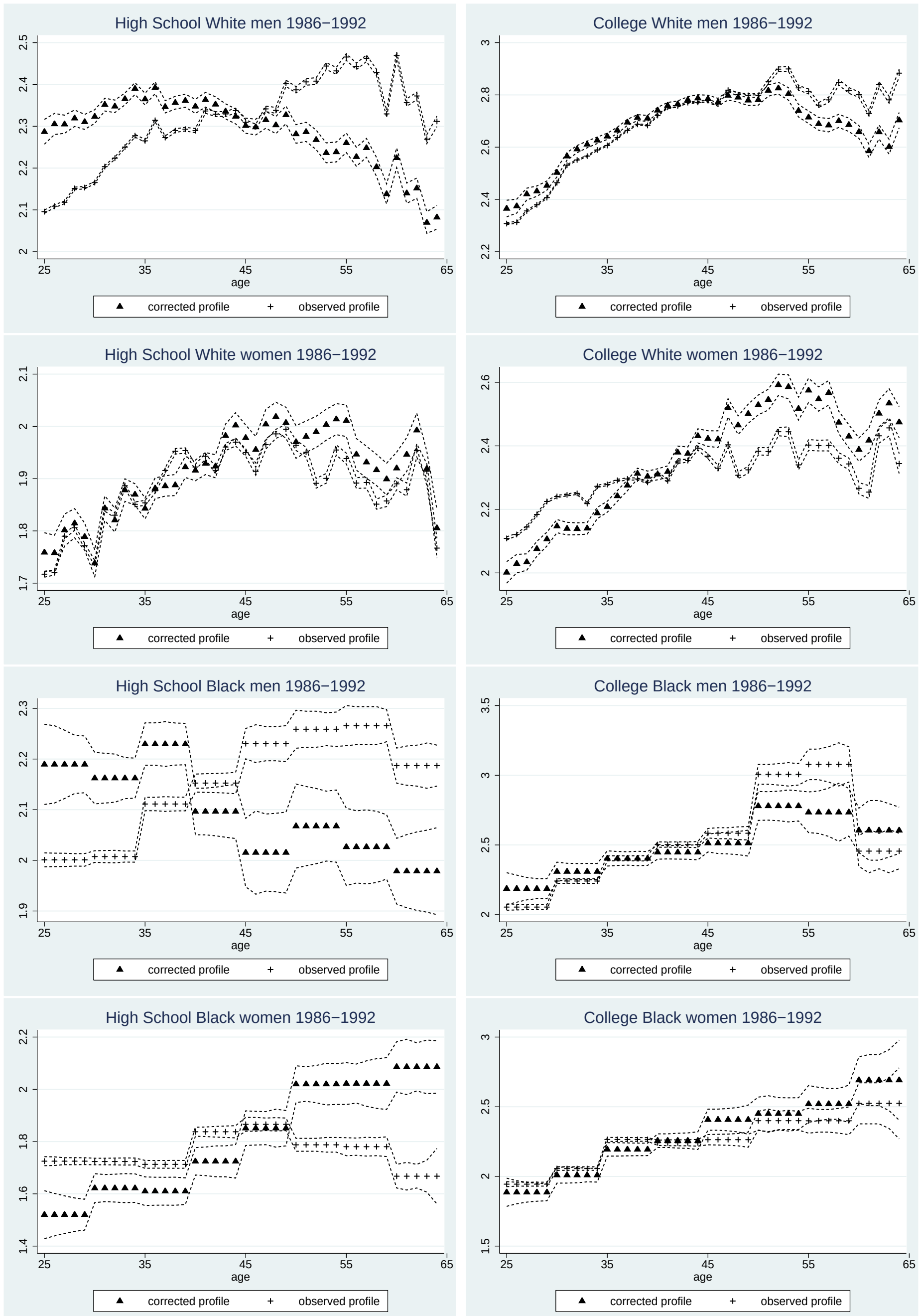


Figure 1.12: Age-earnings profiles 1986-1992

For men, the combination of cohort and selection effects yield corrected profiles which are lower than observed profiles at later ages. While this could be an indication of positive selection into employment, this result seems to be explained almost integrally by strong negative cohort effects, given that the fixed effects of workers and non-workers did not seem to be statistically different at older ages. The reverse is observed for women.

1.4.1 Comparing Age-earnings profiles with the NLSY79

Our paper aims to distinguish the role of cohort effects and selection in the corrected profiles analysed in Section 1.4. For this reason, we repeat the estimation of age-earnings profiles using the National Longitudinal Survey of Youth 1979 (NLSY79). The NLSY is a nationally representative sample of 12,686 young men and young women who where 14 to 22 years of age when they were first interviewed in 1979. We use the NLSY until 2008. The NLSY includes three samples; the main one which is representative of people born between January 1, 1957 and December 31, 1964, totalled 6111 in the first wave; a supplement which tries to oversample from the most representative non-White ethnic groups and poorer groups in the US, mainly Hispanics, Blacks and economically disadvantaged white youth born in the same period; lastly, a sample of the military population which was born in the same period. We exclude the poor and the military supplement as we had excluded the SEO sample for the PSID, and we continue looking at heads of household only.

By following fewer cohorts (all individuals born between 1957 and 1964), we want to confirm how much of the correction we obtain is due to selection into employment and not cohort effects. We chose NLSY79 because it covers most of the lifecycle years we have focussed on. In 1979, the youngest were 14, and by 2008, the oldest were 51. At the same time, this data set is known to have lower attrition rates than the PSID because contrary to the PSID, individuals who attrite one year can return to the sample in later years. According to Fitzgerald et al. (1998), only 52% of the original

1968 PSID children sample were still in the PSID 21 years later and, according to Wu & Li (2005), still 57,6% of the original sample remains until 2004, i.e. 25 years later. Regardless of this, the retention rates were still over 70% for all groups in 2004. Because we are focussing only on individuals between 25 and 64 years old, we only get observations from 1982 until 2008, so that our effective sample period is 20 sampled years spanning 27 years. Because this sample is so young, only 6% of the total sample of heads of households are at least as old as 25. Otherwise, over 62% who started the panel were still interviewed in the last wave available 2008.

Table 1.4 shows some summary statistics of the sample we are using for estimation. We have 5284 individuals available, which is the same order of magnitude of the PSID sample, even though we are looking at a much more homogeneous cohort. The longer time span also guarantees that we lose a very small proportion of individuals in estimation because out of 20 years, more than 90% of individuals from any group have worked at least twice. Inactivity rates are much lower than for the PSID, which is mainly because the NLSY sample is too young (the oldest individuals are 51). Average log real wages are also slightly higher for all groups. Attrition rates, and as discussed, are much lower, which is surprising, given that this is a longer panel.

Table 1.4: Characteristics of the eight labour market groups defined by gender, skill and race: NLSY

	Men				**Women**			
	White		**Black**		**White**		**Black**	
	High School	College	High School	College	High School	College	High School	College
N	1077	1178	208	110	980	1350	185	196
log real hourly wage rates	2.4	2.8	2.1	2.5	2.1	2.5	2.1	2.4
Inactivity rates (%)	7.0	7.0	7.0	7.0	7.0	6.7	8.4	7.1
Attrition rates (%)	14.5	12.2	10.2	12.1	11.6	8.9	12.3	9.7
% individuals lost in estimation	2.5	2.7	2.9	1.8	5.0	2.0	8.1	3.1

While more homogeneous in terms of cohort effects, the NLSY sample includes individuals who experienced significant changes in family structure, in college participation, in labour demand, and in wage inequality. Earnings of earlier ages are thus estimated based on individuals born earlier, and earnings of later ages are estimated based on individuals born later. Fixed effects of mid ages are estimated using the same birth cohorts. Therefore, by narrowing down the number of birth year cohorts being analysed compared to the PSID, we are more likely to see a lower impact of cohort effects relative to selection effects using NLSY.

Our wage regression equation is similar to the one presented in Eq. 1.2, but we include an additional dummy variable to represent the years of the NLSY not included in the PSID. Table 1.5 shows the χ^2 tests of selection and attrition we conducted and explained in Section 1.3.2. For all groups, there is selection both into employment and attrition.

We repeat the analysis of fixed effects of workers and non-workers, and plotted the fixed effects profiles of these two groups, as well as the total sample's fixed effects (Figure 1.13). For high school groups of both sexes and races, workers are better than non-workers, confirming a positive selection into employment. This result was not clear when using the PSID. For College groups, positive selection is not as clear. While it seems to be confirmed for College Black women, it is not at all confirmed for their White counterparts. One reason why this may be the case may be the typical non-labour income faced by these two groups of women. While College White women are more likely to have an employed spouse, that is not the case for College Black women. For College men, while non-workers seem to be lower productivity for most ages, this is not the case for the whole profile. Another difference between these profiles and the profiles estimated using the PSID is the fact that most of the age-fixed effects profiles for the whole sample are closer to zero and flatter, which conforms to our assumption that cohort effects are likely to be less pronounced using the NLSY.

Figure 1.14 further analyses the impact of attrition in these profiles and suggests that attritors tend to be more productive, specially for women. The inclusion of attritors in the profiles led to corrected fixed effects profiles closer to a zero-centered line for some of the groups, even if not for all.

The corrected age earnings profiles using the NLSY, are in Figure 1.15. Results are striking and show that the corrected and observed profiles are practically the same, specially for White groups. While workers and non-workers exhibit different average fixed effects for most ages, the proportion of non-workers is often too small to show substantial deviations in the corrected age-earnings profiles. But the basic pattern of selection persists: the selection into employment of women, specially at later ages, does not seem to be positive. The corrected profiles lie marginally above the observed ones. For men, this is not the case.

Table 1.5: Testing for selective employment and attrition using the NLSY

	$H_0 : \overline{FE}_w - \overline{FE}_{Nw} = 0$		$\overline{FE}_{\text{obs}} - \overline{FE}_{\text{atr}} = 0$		$\overline{FE}_w - \overline{FE}_{\text{exc.w}} = 0$	
	χ^2	***p*-value**	χ^2	***p*-value**	χ^2	***p*-value**
High School White men	582.3	0	304.3	0	442.7	0
College White men	479.8	0	122.1	0	441.2	0
High School White women	309.6	0	238.3	0	908.2	0
College White women	413.5	0	143.7	0	474.4	0
High School Black men	150.8	0	152.7	0	166.7	0
College Black men	179.4	0	201.7	0	205.1	0
High School Black women	129.9	0	195.3	0	340.1	0
College Black women	251.4	0	188.1	0	278.3	0

Results based on the χ^2 test $(R\varphi)\left(R\text{CovR}'\right)^{-1}(R\varphi)' \sim \chi^2_q$, where φ is the vector of estimated fixed effects R is the projection matrix that transforms the fixed effects into a difference in means between two groups Cov is the covariance matrix of the fixed effects and q is the number of restrictions, one for each age level.

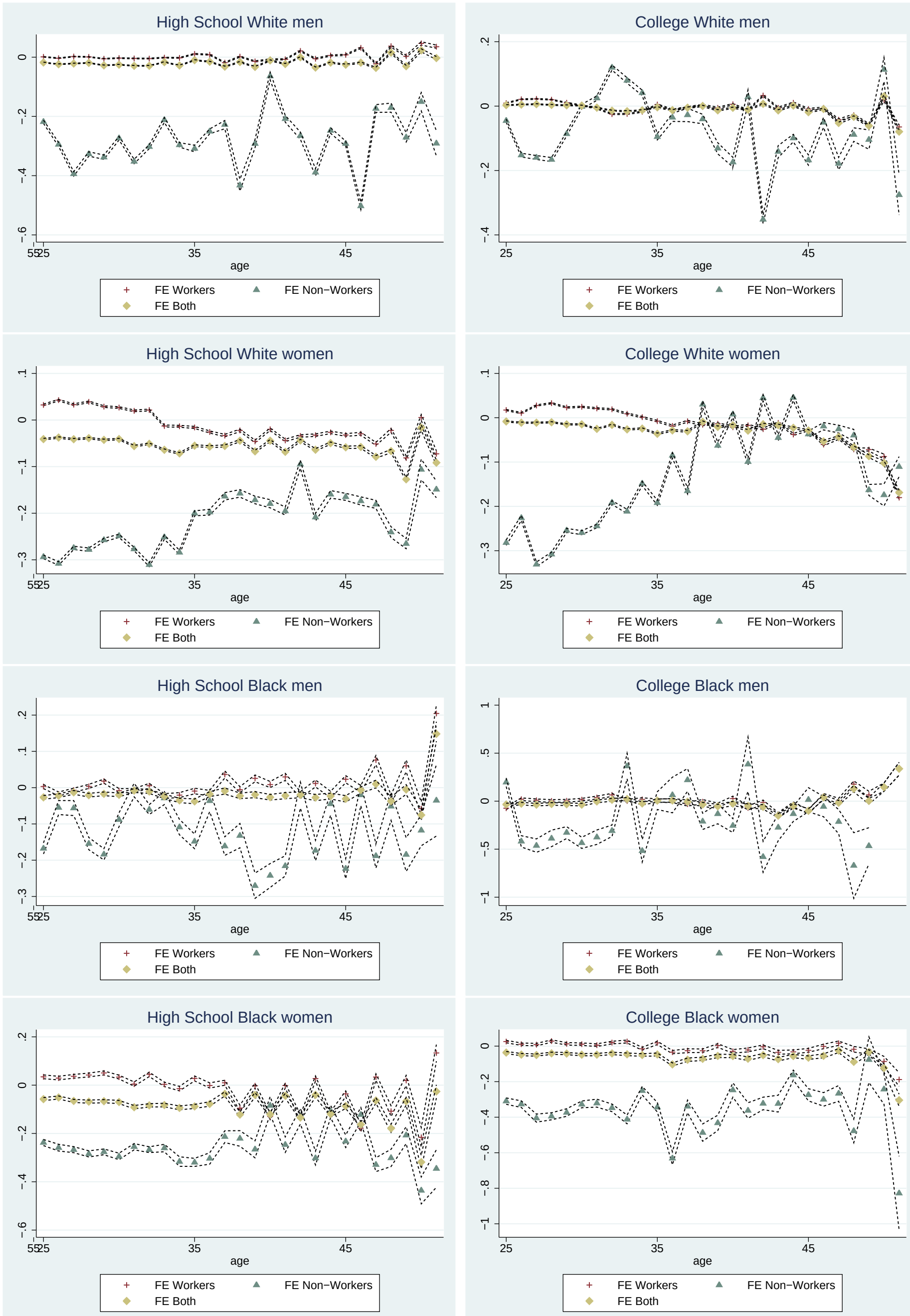


Figure 1.13: How different are workers from non-workers: using the NLSY

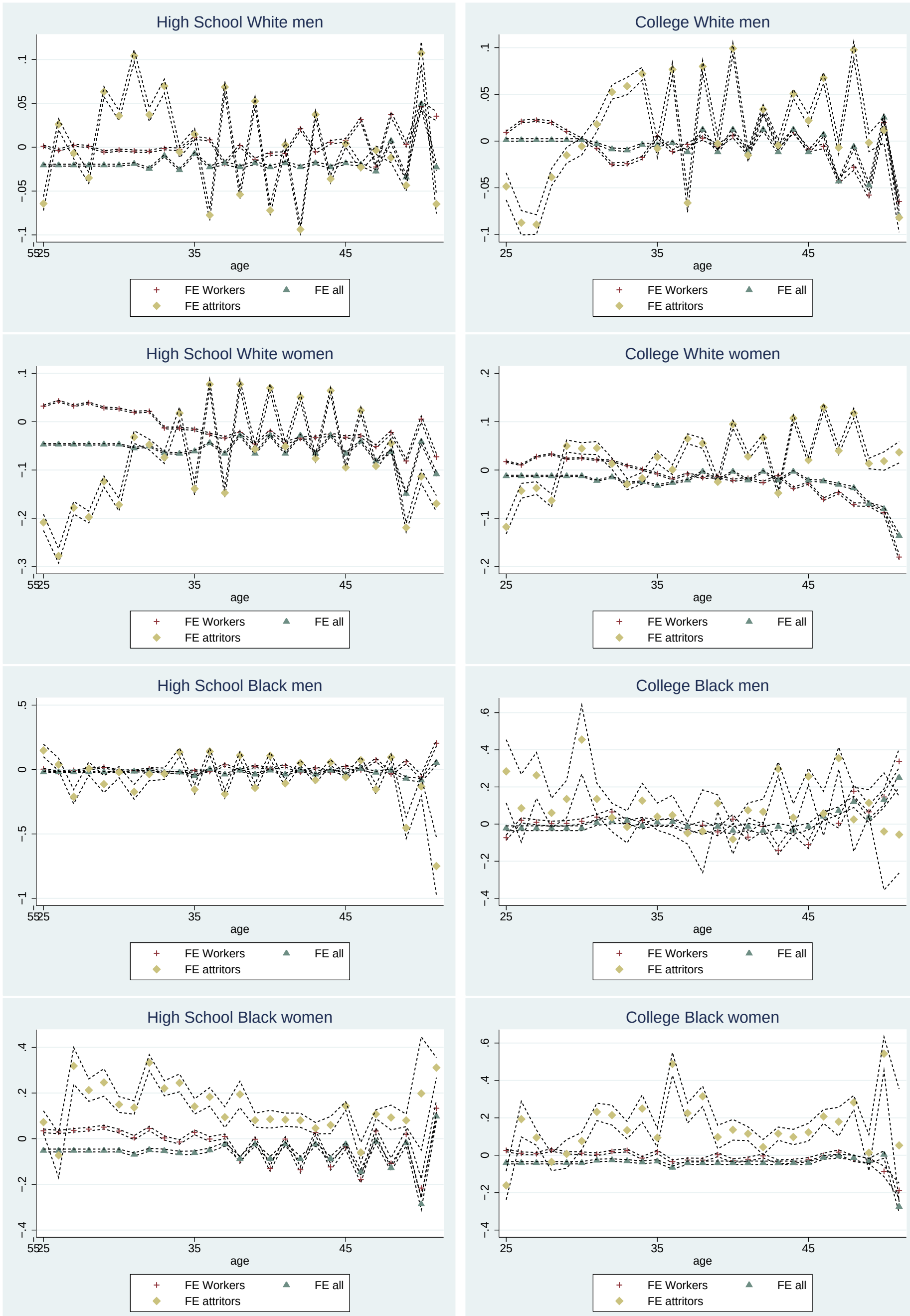


Figure 1.14: How different are attritors from those who stay: using the NLSY

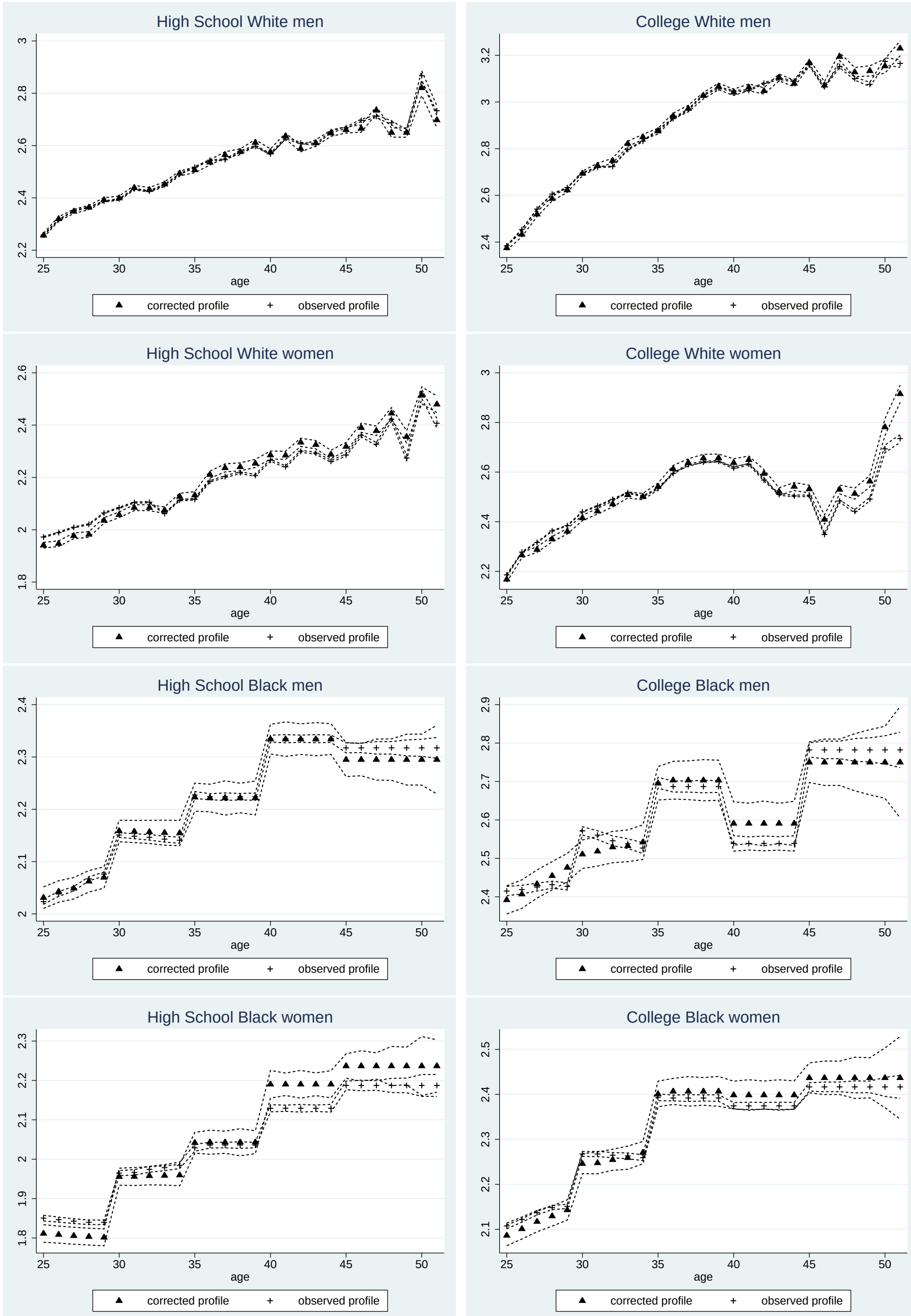


Figure 1.15: Age-earnings using NLSY: following cohorts born between 1957 and 1664

1.4.2 Exogenous wage shocks: a discussion

The key assumption of this model which allows us to impute a potential wage to non-workers is the assumption that wages behave according to Equation 1.2. According to our model, wage growth does not depend on previous experience nor on unobserved heterogeneity. This section uses the NLSY and estimates the fixed effects using a shorter time span (up to age 40) and tests how it correlates with future average hourly wage growth. Wage growth was estimated as a log difference. When NLSY starts with a bi-annual frequency, the log difference was divided by two, assuming equal growth in both years. Table 1.6 shows the p-values of the correlation between the individual fixed effects estimated with shorter time age spans, and annual average wage growth. Results show that for most groups, the correlation between fixed effects and future wage growth is often negative and not statistically significant. Hence, despite the strong assumptions made to estimate the wages of non-workers, data seems to be consistent with the assumptions made.

Table 1.6: Testing for exogeneity of shocks to wage growth using the NLSY

	Correlation coefficient	p-value
High School White men	-0.0031	0.9290
College White men	-0.0052	0.8727
High School White women	-0.0445	0.2326
College White women	0.0025	0.9346
High School Black men	0.1577	0.0539
College Black men	0.2111	0.0511
High School Black women	-0.0289	0.7496
College Black women	-0.1890	0.0193

1.5 Who works and why? A life cycle model of labour supply

The previous sections show that workers of different groups defined according to gender, race and skill, seem to have different incentives to participate in the labour market, at different moments in the lifecycle. While College White women at later ages seem to make their labour supply and retirement decisions based on factors beyond their unobserved productivity, this does not seem to be the case for men or College Black women. While this may well be explained by differences in non-labour income across these groups, initial wage level, and wage growth over the lifecycle also seem important indicators of labour supply decisions and, as the fixed effects estimation results of our wage regressions showed in Table 1.2, these vary substantially across these groups too. This section will explore models of intertemporal labour supply decisions proposed in the literature. It will discuss the extent to which models are able to replicate the decreasing participation patterns observed in the data, as well as the differences across these groups in terms of non-labour income, wage level and growth, and how these factors seem to explain why selection into employment varies across women, skill group and race.

The initial models of intertemporal labour supply were discussed in Heckman (1974a). When the incentive to save (the interest rate) equals the discount rate, flexible hours of work always leads workers to work more when wages are higher (and higher earners to work more than lower earners), which results in a profile of hours of work tracking the wage schedule. The consumption profile will however depend on how the marginal utility of consumption changes with leisure. If leisure and consumption are substitutes, then individuals find it optimal to consume more when wages are higher. If leisure and consumption are complements, then consumption will decrease with wages. When the discount rate is different from the interest rate, the intertemporal substitution effect where there is tracking between wages and hours of work remains, but may be offset by very large incentives to accumulate wealth (often unreasonably high given common parameterisations of such a life cycle model), specially if consumption and leisure are substitutes. These models are deterministic and often have a closed form solution. Low (2005) discusses an extension of this model which includes uncertainty in wages. We will show how this simple extension allows us to explain our results to a great extent. We also simulate a deterministic model with endogenous wages, given the strong assumption we have used in this paper of exogenous shocks to wages, but this model does not seem to offer a better explanation to our results than the model with uncertainty and a simple wage formation process.

The baseline optimisation model is in Eq. 1.4.

$$\max_{\{c_s,l_s\}} \sum_{s=t}^{T} \beta^{s-t} \frac{\left(c_s^{\eta} l_s^{1-\eta}\right)^{1-\gamma}}{1-\gamma} \tag{1.4}$$

subject to

$$\begin{aligned} A_{s+1} &= (1+r)\left(A_s + (H - l_s)\, w_s - c_s\right) \\ A_{T+1} &= 0 \\ H \geq l_s &\geq 0 \end{aligned}$$

The utility function chosen for each period s is an isoelastic Cobb Douglas and individuals choose consumption c and leisure l in each period which maximise the discounted lifetime utility. This is subject to a wealth accumulation law of motion, where wealth A increases deterministically according to an initial wealth level, and the accumulated amount of income saved or borrowed throughout life. There are no borrowing constraints. In any given period, individuals increase their stock of wealth when their consumption is lower than their earnings, and eat away their wealth otherwise. There is also a no bequest constraint, and a restriction on the amounts of leisure available in each year, which have to be nonnegative and never higher than the total amount of hours allocated for leisure and work H. The first order conditions yield the optimal intratemporal leisure decision as a function of current consumption.

$$l_t = \max\left\{\min\left\{\frac{c_t}{w_t}\frac{1-\eta}{\eta}, H\right\}, 0\right\} \tag{1.5}$$

This means that in effect there is only one control variable to account for in the dynamic programming process, which we find by finding the value on consumption that maximises the value function over the grid of wealth levels, for each wage rate. The results of this simple model, which assumes a deterministic exogenous wage process, is presented next. The following two sections present the results of two models which change the wage determination process, and assess how closely these extensions produce simulated profiles similar to observed profiles.

1.5.1 The basic lifecycle model with flexible labour supply

Wages are assumed to grow at an exogenous rate b for the first 30 years of work and decrease by the same factor from then on until the end of the working life set at $T = 40$ years.

$$\begin{cases} w_t = w_0 \times b^t & \text{if} \quad t \leq T-10 \\ w_t = w_0 \times b^{(T-10)+(T-10-(t-1))} & \text{if} \quad t > T-10 \end{cases} \tag{1.6}$$

This model will be simulated for individuals who vary according to their initial wealth level (a proxy for non-labour income) and the wage profile they face (with both initial value and growth rate varying). We have used the same initial wealth for all individuals. We use 200 equally spaced values, from 0 to to 100 tens of thousands of US dollars (deflated by the CPI 1985 index). This grid was adjusted each period according to the maximum savings and debt that an individual could accrue in the previous period. For the initial wage levels, we use the deciles of the average *observed* wage distribution faced by each of the eight groups discussed in our paper (see Figure 1.3 for a diagram of the average wage profile conditional on age of these eight groups). For the wage growth rates, we use the average growth of each of these profiles over the entire life span[6].

[6] All wages were left censored at log(3.5). The average wage growth rate observed for low educated White men was 0.00878; for high educated White men, it was 0.00146; for low educated White women: 0.01098; for high educated White women: -0.00762 (for this group, we assumed a monotonically decreasing wage profile); for low educated Black men: 0.03401; for high educated

The remaining parameters of the model have been chosen based on Low (2005). We set the interest rate to $r = 0.016$ and the discount factor to $\beta = 1/1.009$. We chose a neutral leisure share $\eta = 0.5$, but kept $\gamma = 2.2$ and $H = 5200$. These parameter values assume a slight incentive to save and substitutability between consumption and leisure.

We solve for this model using backward induction. Consumption and leisure are treated as continuous variables. Because leisure is a limited variable, we solve for three different optimisation problems for each wealth level - the continuous case, where leisure is defined as a function of consumption as in Eq. 1.5, the case of maximum leisure (the inactivity case), and the case of minimum leisure[7]. We choose the consumption and leisure which yield the maximum value for the value function. At period T, optimal consumption and leisure are found analytically using the terminal condition. For each wealth level, optimal consumption equals the sum of current wealth and the earnings obtained in that period. By comparing the value for the utility function across all three possible scenarios for leisure (and its resulting consumption) we obtain the optimal consumption and leisure for each value of wealth with which the individual starts the last period. For all the remaining periods, we find the policy function of consumption and leisure by maximising the value function for each value of wealth available at the beginning of each period. We have used linear interpolation to match the wealth (and the value function) resulting from an optimal choice to the wealth (and value function) values available in the grid found in the previous step.

Figures 1.16, 1.17 and 1.18 show the simulated profiles for all eight groups of participation, hours of work and wages (both observed and corrected) resulting from this model. A detailed analysis of the consumption, leisure and wealth profiles, and

Black men: 0.01033; for low educated Black females: 0.01759; for high educated Black females: 0.00666.

[7]Each period's utility function in Eq. 1.4 is not well-defined when either consumption or leisure are close to zero, given our choice of parameter values. We have set a minimum for either consumption or leisure of 1.

of how these vary with wage and wealth parameters, is available in Appendix D. These profiles were drawn for two different levels of initial wealth. Participation increases with age for all groups except for College White women, the group which faces negative wage growth. Participation is not monotonically increasing for College Black women either, who face a relatively low wage growth rate and low startup wage. So in effect, results show that participation tracks wages and tends to be higher when wages are higher or growing positively. This result is confirmed in Figure 1.17, which shows hours of work displaying the same shape as the wage process that generates them. The only exception is the College White male group, where hours of work continuously decrease over the life course, decreasing at a slightly faster rate when wage growth becomes negative. This is the case because wage levels are so high (relative to their growth rate) that allows this group to sustain increasing consumption and leisure profiles while accumulating wealth early in life, which is run down at an increasing rate when work-related earnings start decreasing. For all groups, higher non-labour income reduces participation in the labour market.

The resulting wage profiles show that wages of those who work are, if different, always higher than unconditional wages, which suggests positive selection into employment. This prediction of the model proposed in Heckman (1974a) has motivated most authors (e.g Petrongolo & Olivetti, 2005; Neal, 2004; Heckman et al., 2000; Chandra, 2003; Blundell et al., 2007) to assume positive selection into employment instead of being silent about the relative or absolute productivity of non-workers. While this paper has shown that this may not always be the case, this model cannot explain the sharp decrease in participation observed in real data (nor the constant slight decrease in hours of work shown in Appendix A). Our two next models propose simple extensions to the wage process which lead to simulated profiles closer to the data.

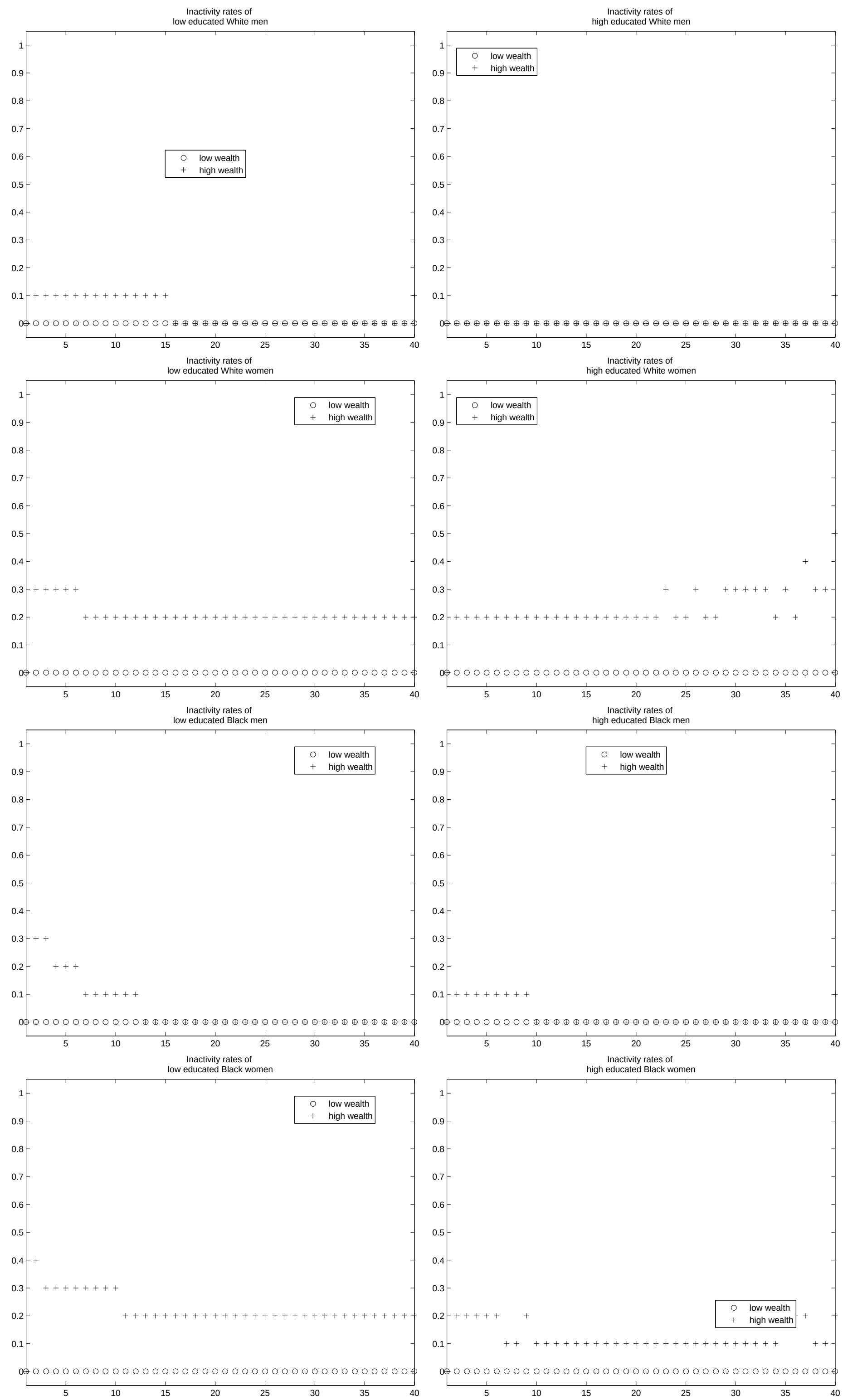


Figure 1.16: Participation rates in the basic model

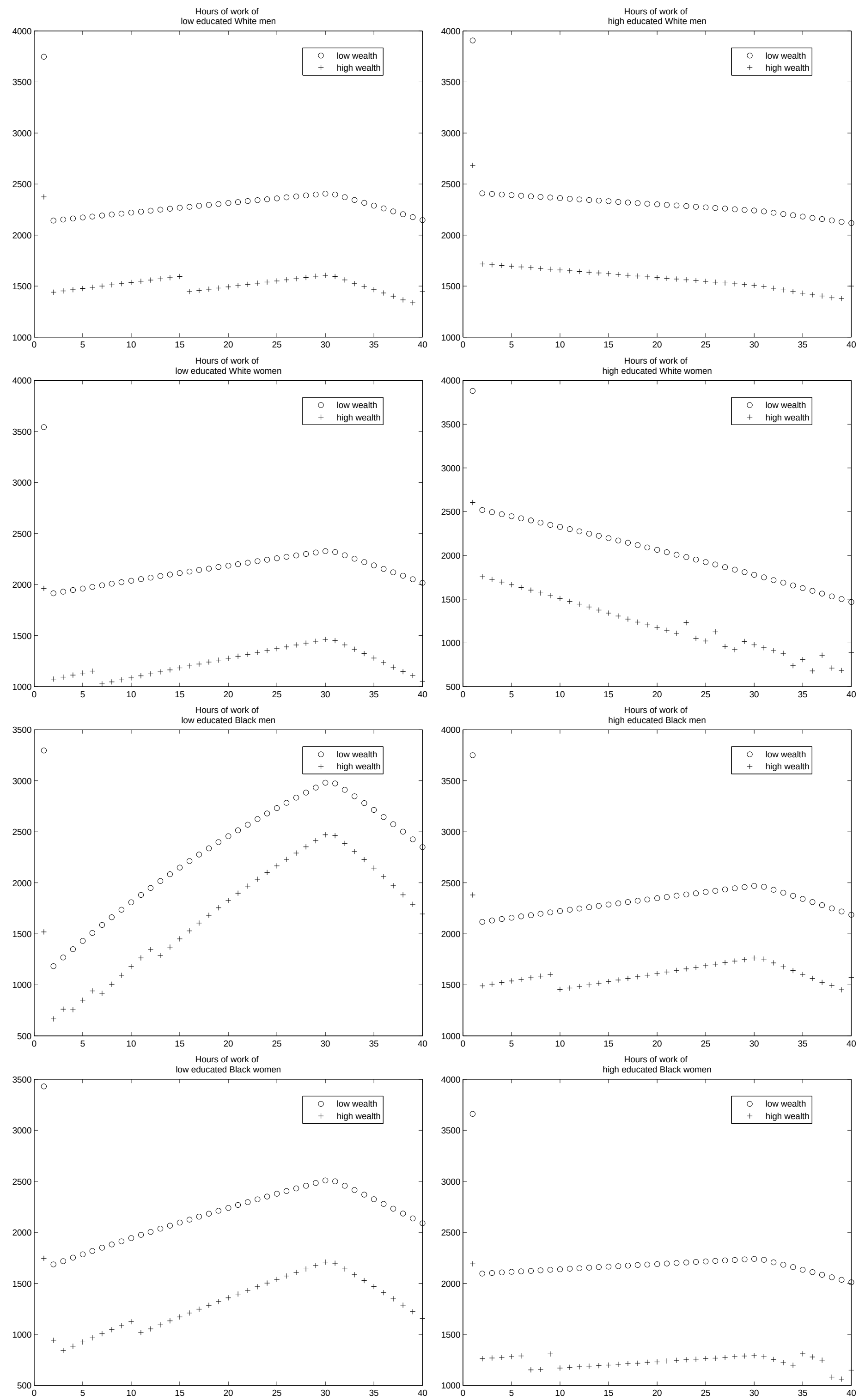


Figure 1.17: Hours of work in the basic model

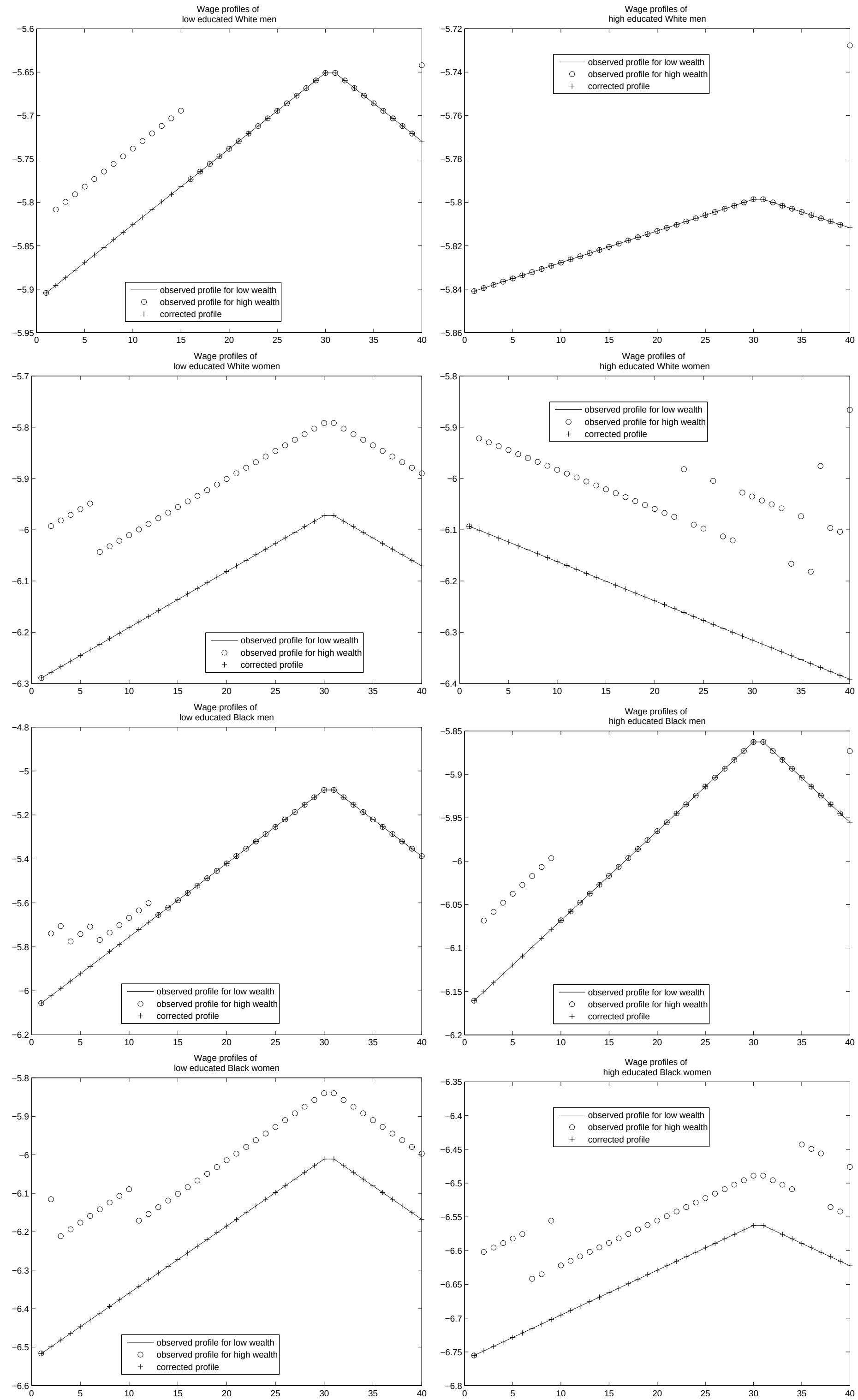


Figure 1.18: Wage profiles in the basic model

1.5.2 The lifecycle model with flexible labour supply and deterministic endogenous wages

Several authors have argued that when labor supply decisions not only determine current, but also future wages, this investment effect of participation leads individuals to work longer hours earlier in life (see e.g. Eckstein & Wolpin, 1989). This literature, often focussing on the participation choices of married women, can explain why participation may be high at earlier ages. In the standard model discussed in the previous section, individual wages grow at an exogenous rate, whether or not they have worked in previous periods. It does not take into account human capital accumulation acquired on the job, nor its depreciation. This section shows a very simple deterministic model with endogenous wage growth, where wages in each period t depend on the total number of years the individual has worked before, call these e. Each year working augments wages by a factor b and each year away from the labour market depreciates wages by the same factor. The wage determination process is defined in Equation 1.7.

$$\begin{cases} w_t = w_0 \times b^{1-t+\mathrm{I}(t=1)} & \text{if} \quad e = 0 \\ w_t = w_0 \times b^{3-t-\mathrm{I}(t=1)} & \text{if} \quad e = 1 \\ \begin{cases} w_t = w_0 \times b^{t-1} & \text{if} \quad t \leq e \\ w_t = w_{t-1}/b^{t-(e+1)} & \text{if} \quad t > e \end{cases} & \text{if} \quad e \geq 2 \end{cases} \tag{1.7}$$

In this model, labor supply decisions not only impact on current income and wealth, but also on future wages and income possibility sets. This presents an incentive to work more at earlier ages and to retire earlier when the returns to the investment in human capital are low - the investment effect. However, because the returns of labor supply are now larger than in a model with exogenous wage growth, and because the cost of accumulating debt has become even cheaper, endogenous wage growth may lead to more consumption and more leisure at earlier ages because these can be financed in a shorter spell of time - the consumption effect. This will be even more so the higher the wage growth rate faced by individuals. The investment effect is similar to an increase in r relative to β which individuals can now influence with their labour supply decisions.

The parameters of this model are the same as for the basic model, except that we have not considered negative wage growth, instead looked at what happens when there is no wage growth at all (benchmark situation where wages do not depend on accumulated wealth). Figures 1.19, 1.20 and 1.21 show the simulated profiles of participation, hours of work and wages for all eight groups. Appendix D describes the results of our simulations in detail. Results are very similar to the ones in the previous section, when wages were assumed to be exogenous. Leisure continues to be decreasing for most groups, and inactive people, if any, continue to be the low earners with high initial wealth, whose incentive to work in the start of their life is low and who consume their wealth at a constant rate until the end. The groups who face an increasing number of hours of leisure over time, and therefore decreasing hours of work as Figure 1.20 shows, are College White men, who have the highest wage levels and a very low wage growth rate, and College White women, who face no wage growth during their lifecourse and only have moderate wage levels. These are also the two groups who do not find it optimal to incur any debt because recouping debt by working is more difficult than for other groups, whose wage growth rates, and therefore, whose returns to experience, are much larger. These two groups find it optimal to increase both leisure and consumption over the lifecourse, even if the growth rate of either is more modest than the growth rate of consumption for other groups with higher returns to experience.

The wage profiles resulting from these optimal decisions are shown in Figure 1.21. As in the previous section, selection into employment, if individuals do stay out of work which the model can hardly produce, is positive and occurs earlier in life. Endogenous wages do offer an additional theoretical effect on intertemporal choices that can offset the intertemporal effect found in Heckman (1974a), but there are still very dominating discounting and intertemporal effects making leisure higher at earlier ages than in the data, specially for moderate and high wage growth.

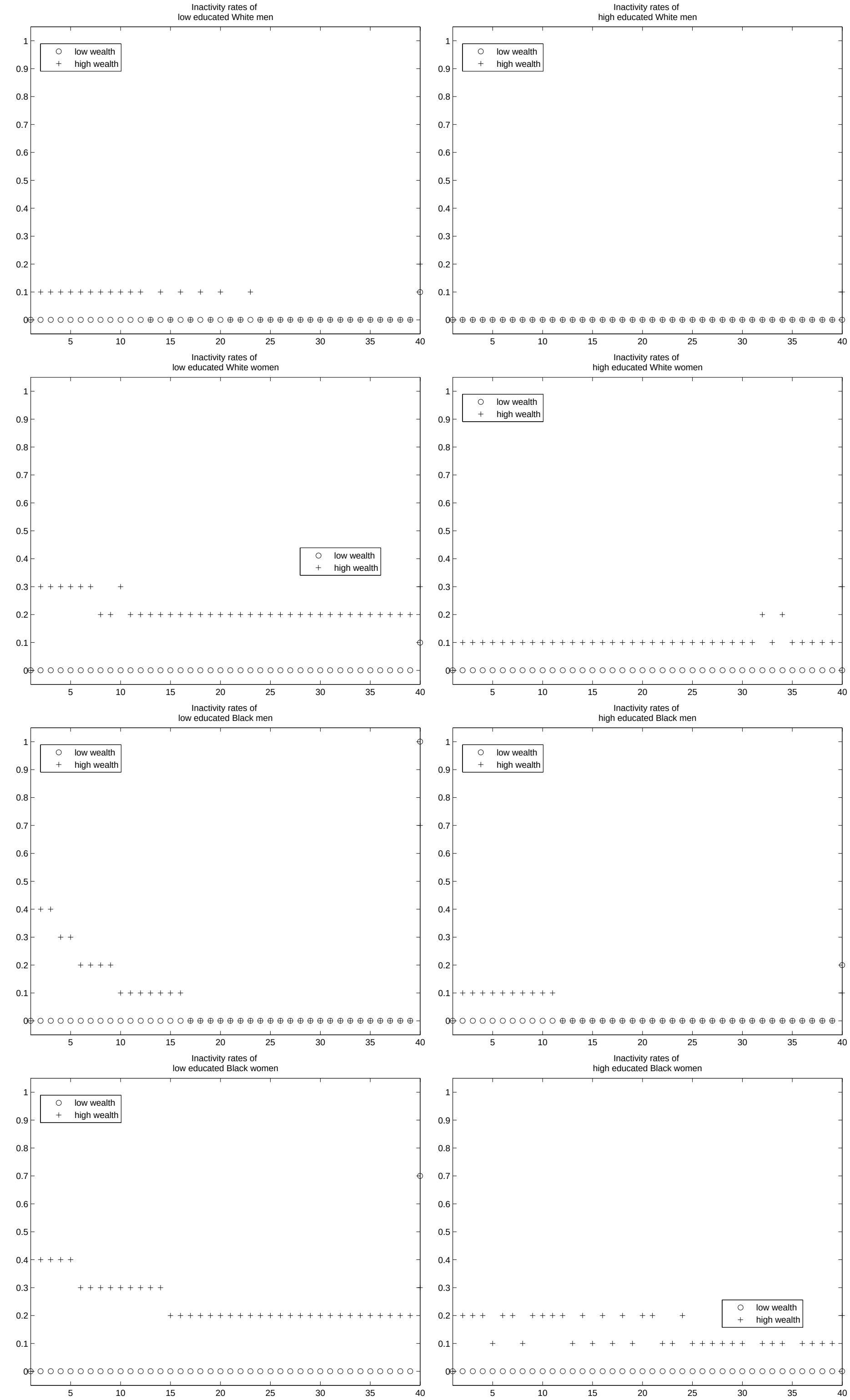


Figure 1.19: Participation rates in the model with deterministic endogenous wages

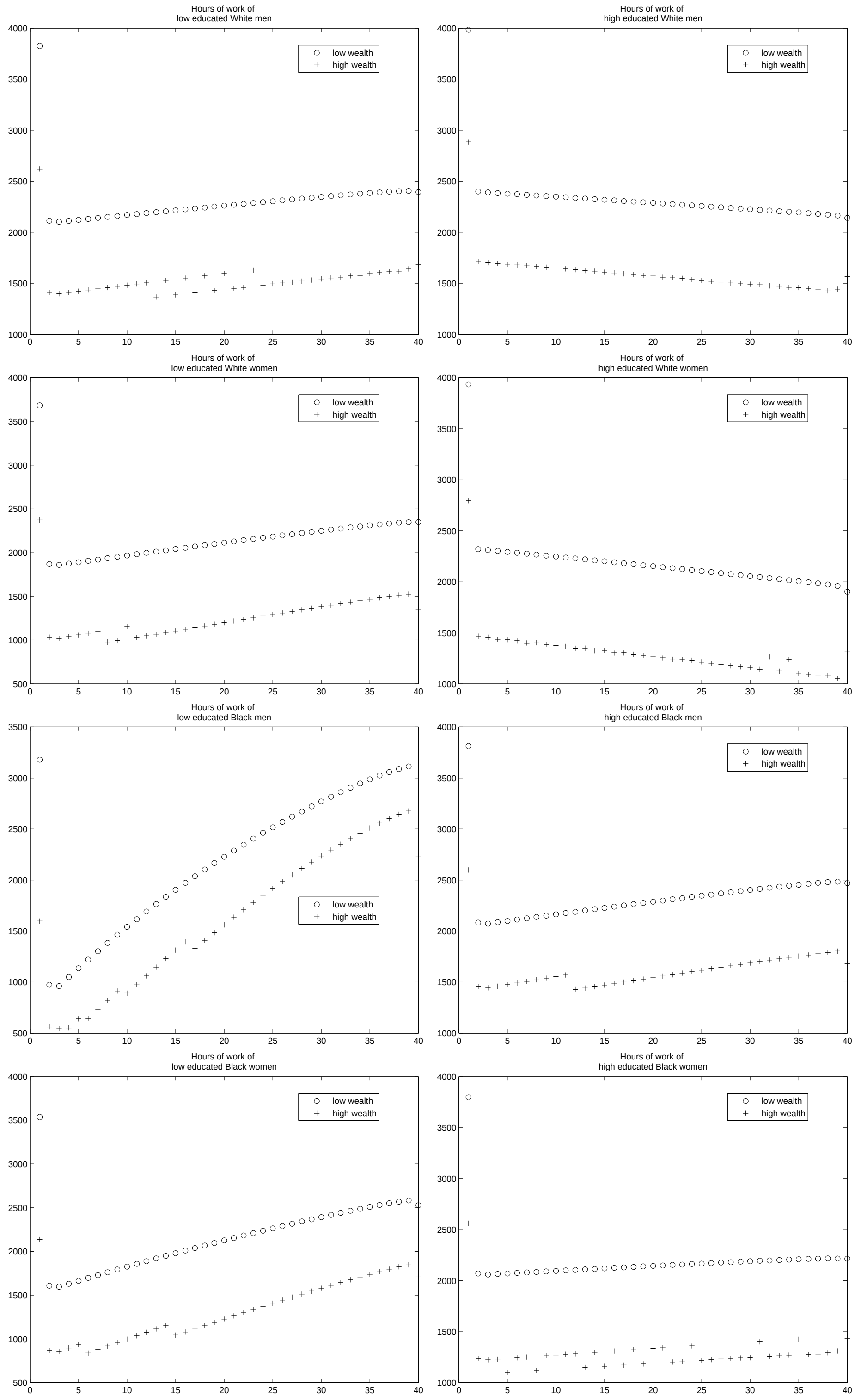


Figure 1.20: Hours of work in the model with deterministic endogenous wages

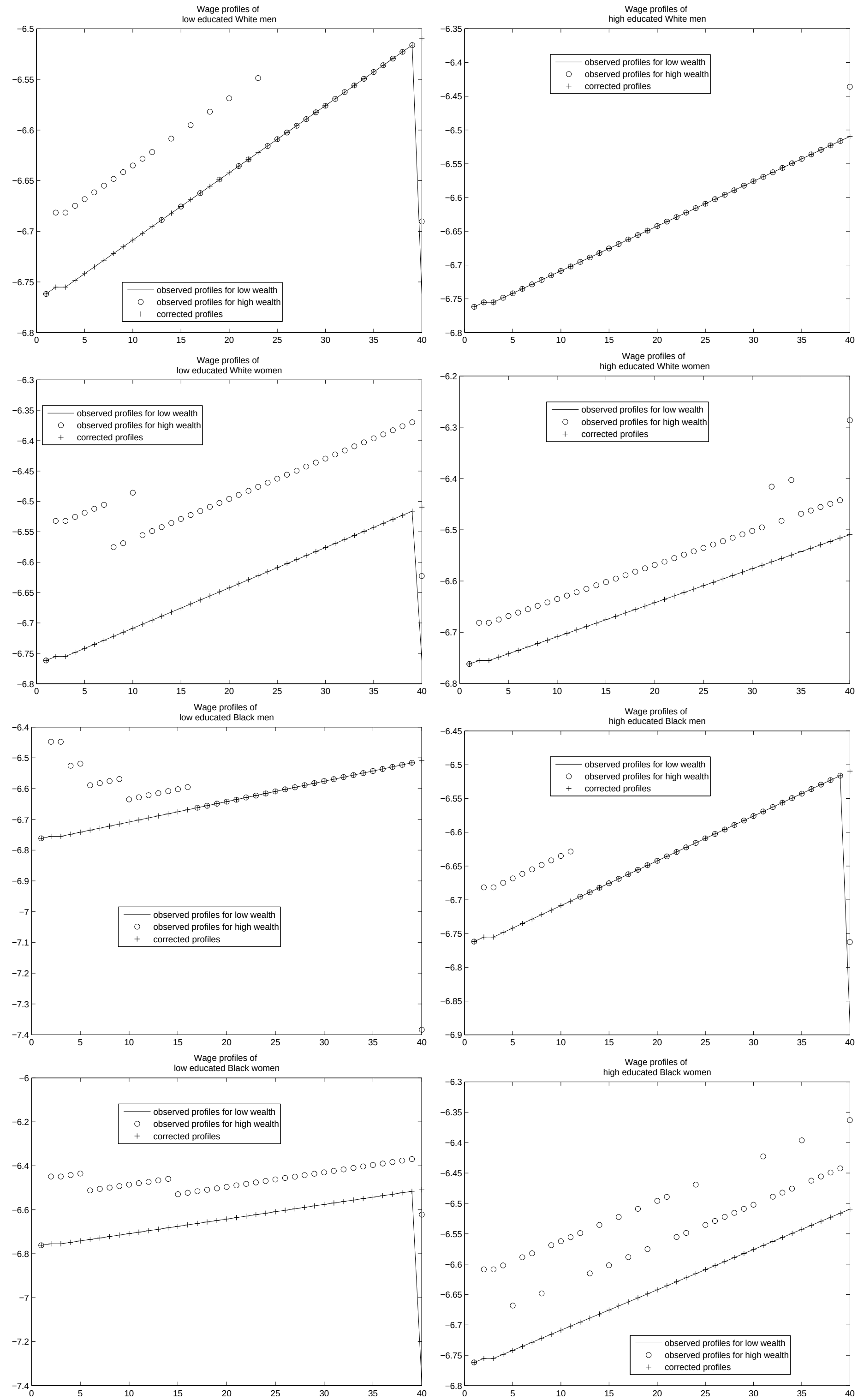


Figure 1.21: Wage profiles in the model with deterministic endogenous wages

1.5.3 The lifecycle model with flexible labour supply and stochastic exogenous wages

Low (2005) shows uncertainty in wages causes individuals to work longer and to consume less at earlier ages in order to build precautionary savings that they can use against future shocks to wages. This buffer stock of wealth will allow individuals to reduce their working hours at later ages, when uncertainty is resolved. We extend our basic model of intertemporal labour supply and add uncertainty to the wage formation process. Wages are subject to a shock drawn from a discrete distribution.

$$\begin{cases} w_t = w_t - 1\epsilon \forall \epsilon \in (0.4, 0.9, 0.99, 1, 1.1, 1.3), \forall T \leq 30 \\ w_t = w_t - 1\epsilon \forall \epsilon \in (0.4, 0.9, 0.95, 1, 1.1, 1.2), \forall T \geq 31 \end{cases} \tag{1.8}$$

To mimic the inverted-V shaped wage profile of the previous sections, we use a probability distribution such that the expected value of the wage is larger than current wage up to period $T-10$, and lower than the wage from then on[8]. The shock is realised at the end of each period, so that once decisions are made in each period, there is a shock to the wealth carried over to the next period. We have simulated this model by generating 5000 wage trajectories over the lifecourse for each group, level of wages and wealth. The optimal choices made under these 5000 scenarios were averaged to produce the resulting profiles.

Figures 1.22, 1.23 and 1.24 show the results. Appendix D discusses these results in more detail. Results are very encouraging. This model replicates decreasing participation over the lifecycle for all groups. The participation rate is higher than in deterministic models due to uncertainty, and leads to very little variation in participation rates across groups or non-labour income levels. We observe a decrease in participation for all White groups except the High School women, and also for Black High school men, but only for a few years. Several results emerge from these figures (as well as figures in the appendix): those with high non-labour income

[8]The probability distribution is the same in both periods $(0.05, 0.1, 0.2, 0.35, 0.2, 0.1)$.

(such as the median College White woman according to Neal, 2004), work less; those with a high wage growth and low initial wage level smooth consumption and leisure over their lifecycle, and run down their wealth almost monotonically. For these groups, those at the top of the distribution in terms of initial wage can work more at all ages, which explains positive selection observed in the male groups. Those with lower wage growth rates (flatter profiles) are therefore more likely to postpone leisure and consumption to accumulate wealth and buffer against future negative wealth shocks. In these groups, those with higher initial wage level will work harder than everybody else in the beginning of their worklife, but retire earlier. Lower productivity workers in these groups work less than everybody else, except at later ages when highest productivity group overtakes them. This means that those who retire earlier can be both high productivity workers, who accumulated enough wealth over the lifecycle and income effects kick in, or low productivity workers, because of lower opportunity costs of staying at home. The former group of high productivity/early retirement has however been overlooked in most of the empirical literature accounting for selection into employment.

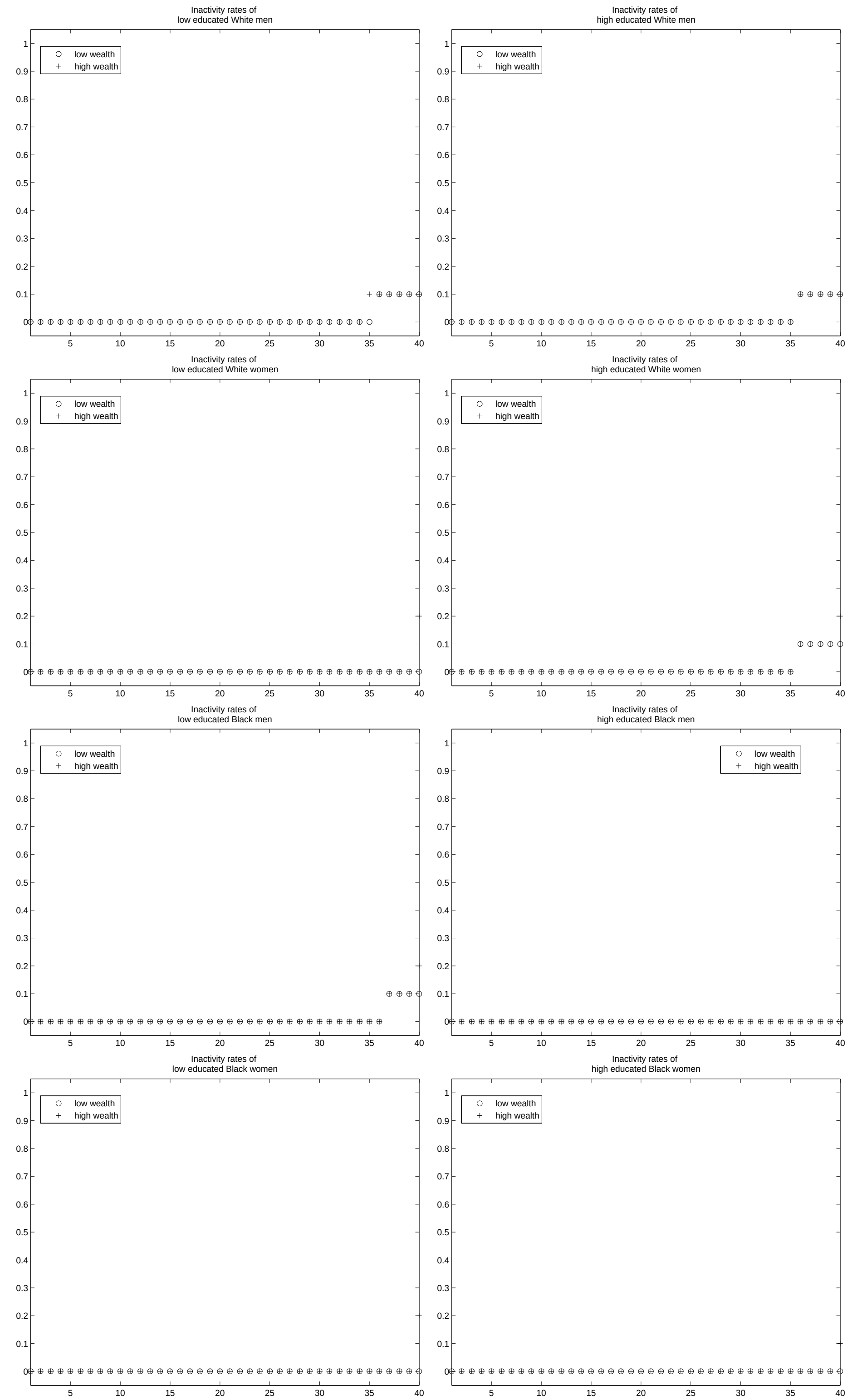


Figure 1.22: Participation rates in the model with uncertainty in wages

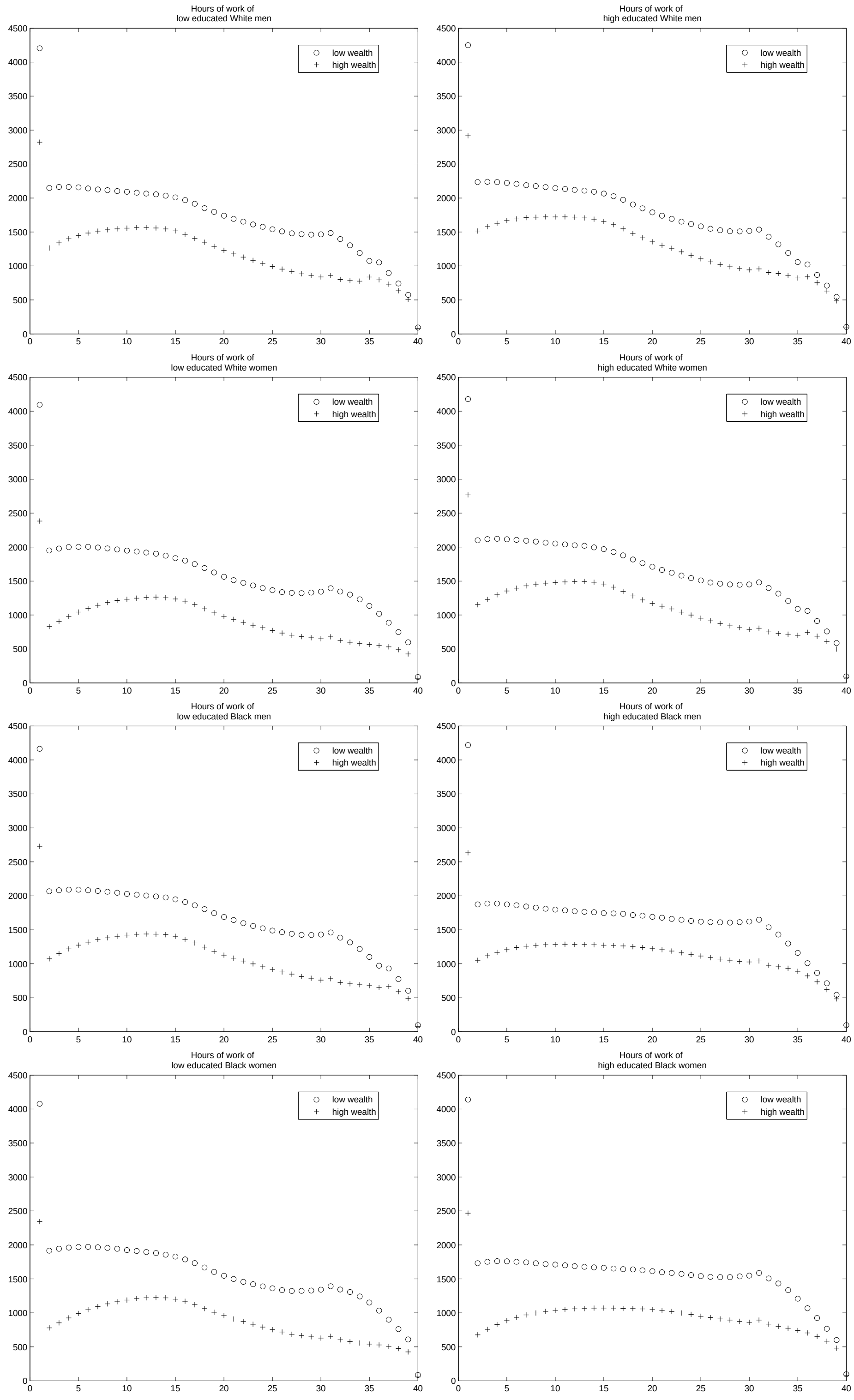


Figure 1.23: Hours of work in the model with uncertainty in wages

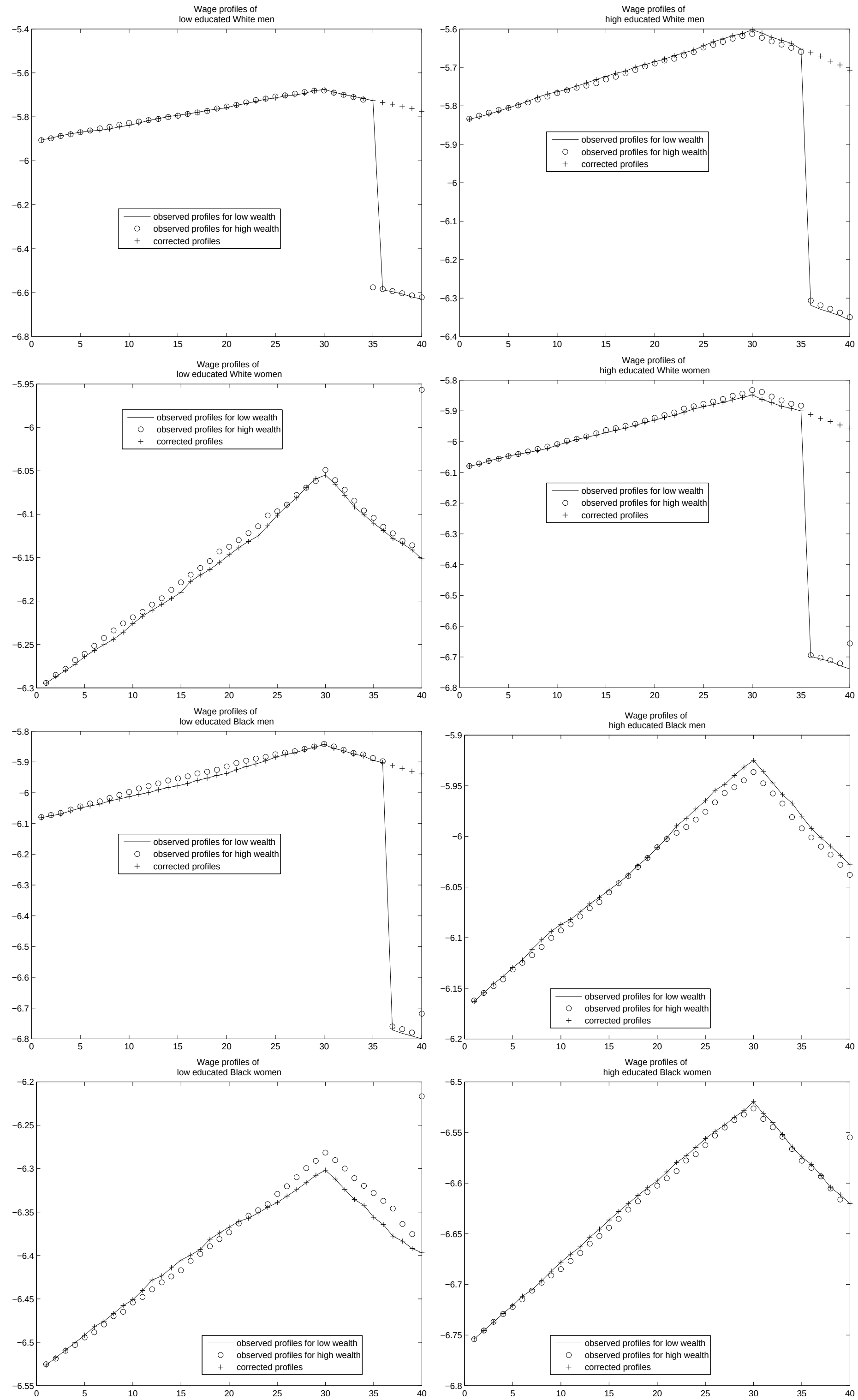


Figure 1.24: Wage profiles in the model with uncertainty in wages

1.6 Conclusion

Age-earnings profiles are instrumental in the analysis of inequality between different social groups (e.g. gender or ethnicity). They are also important in the analysis of intertemporal labor supply decision processes. However, age-earnings profiles constructed solely from averaging observed wages at different ages are contaminated by selection into employment bias. Most of the literature corrects for selection by assuming non-workers are lower productivity than workers. This paper uses a very simple approach that is silent about the relative or absolute productivity of non-workers. We construct potential wages by using estimated fixed-effects from a regression of earnings on age and time-related variables. These are available for everybody, whether they work or even whether they are observed at all. This allows us to build a dataset that adds potential wages for non-workers - correcting for selection. Corrected profiles do not always lie below observed profiles, running counter to the assumption of positive selection into employment. We repeated our analysis using the NLSY to disentangle cohort and selection effects. These results suggest that selection is statistically significant, but workers can be lower earners than non-workers.

These results are still preliminary and several further steps need to be taken for a clearer picture of labor supply decisions. As yet, we have nothing to say about those individuals that have never been observed working. This approach can however be easily used jointly with existing (better informed) procedures in the literature that do try to estimate the profile for all individuals, even those for whom there is no wage information. This is more important for women, the group with lower participation rates, and for Blacks.

Chapter 2

Revisiting the age-happiness profile: Estimating age, period and cohort effects.

Abstract

This paper estimates age-happiness profiles using alternative specifications for age, period and cohort. It discusses the two main methods, fixed effects and constrained generalised linear models, which are used to identify age effects in the happiness literature. This paper will estimate and replicate the findings of previous studies which have used restrictions on the coefficients for age, period and cohort. This paper also proposes an alternative way of identifying the effects of age, period and cohort. Instead of imposing restrictions on the vector of parameters, it explores the discrete nature of the data and redefines age so that age, period and cohort effects can be estimated, even at the individual level. It relies on the fact that not all individuals are born/interviewed on the same day, which creates an exogenous source of age variation within the same birth year cohort. Once linear effects of age, period and cohort are accounted for this way, and once fixed effects can separately identify age and period effects, age-happiness profiles estimated using OLS, fixed effects or ordered probit fixed effects differ from those already found in the literature.

JEL classification: D69, D84, I30

Keywords: age-happiness profile, APC models, linear effects

2.1 Introduction

Figure 2.1 shows age-happiness profiles estimated for different cohorts, and also for the overall sample. The unconditional profile averages the happiness of all individuals of a given age, without accounting for year nor cohort. While this profile seems to suggest happiness decreases with age up to the early 50's, at which point happiness stagnates or even starts increasing again, the profiles observed when following each cohort seem to tell a different story. Conditional on age, we can see that average happiness of cohorts born earlier is higher. This could also be because of time effects being different in the years when different cohorts had the same age. These profiles were estimated using observations collected between 1984 and 2003. The cohort specific profiles of cohorts born between 1939 and 1959 are broadly horizontal shifts of each other, but this is not the case for profiles of older and younger people. In fact, the two most recent cohorts show a flat profile, which sits close to the unconditional profile, and are very distinct from all others. With a quick visual inspection of Figure 2.1, we therefore see that time effects, cohort effects and age effects, all seem to matter in describing happiness.

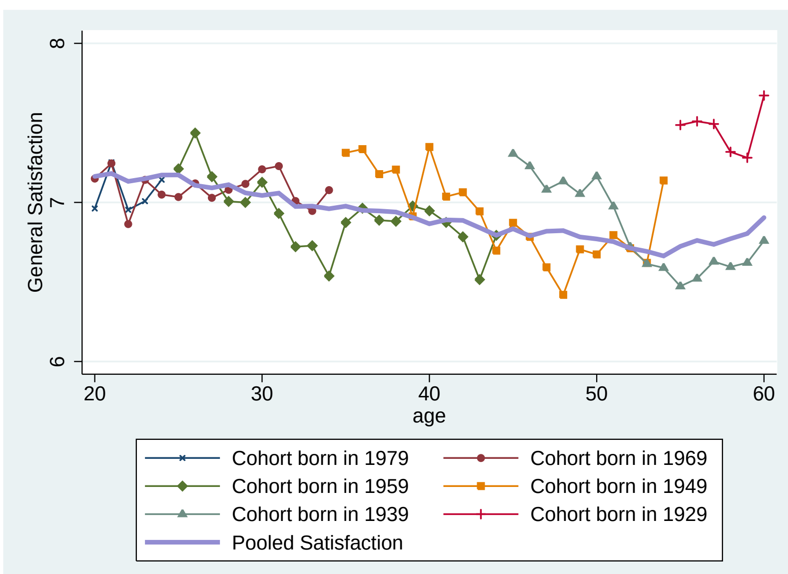


Figure 2.1: The happiness profile in age, following different cohorts 1984-2003

However, because of the linear dependence between age, period and cohort, often happiness equations are estimated without accounting for all three factors. In particular, age-happiness profiles are important per se. Longevity is increasing and it is important to evaluate how happy older people are likely to be. Wilson (1967) concluded that younger people are happier. Most studies in Psychology often find that age has no impact on happiness at all, which is consistent with the hedonic treadmill theory. The Economics literature has often produced a U-shaped age-happiness profile, where the dip is around the age 50, which is consistent with Figure 2.1 (for a review of the literature on age-happiness profiles, see Frijters & Beatton, 2008; Clark, 2002). Easterlin (2001) suggests that this pattern reflects unfulfilled overoptimistic expectations of the young, who adapt to present circumstances later in life. Frijters & Beatton (2008) and Clark (2002) suggest that cohort effects may underlie the relation between age and happiness. While Frijters & Beatton (2008) argues that cohorts are "just a missing aggregate variable specific to an age-group but where we do not know what the missing variable is", other authors recognise that cohort effects are the true essence of social change (e.g. Yang et al., 2008; Cribier, 2005). The author of this paper tends to agree more with the second view, where age effects capture lifecycle regularities we observe across time (and actually not just the cumulative effect of life events which tend to happen at particular stages of one's life), cohort effects capture the evolving social context whose impact affect individuals in different stages of their lifecycle differently. Identifying both effects, and separably from each other, is then a key aspect of research in social sciences.

Age and cohort effects are difficult to account for when time effects also exist, due to the linear dependence between the three variables. A lot of work has been done in Epidemiology, Demography and Sociology to analyse such models. In these areas, the two most common approaches are constrained generalised linear models (CGLIM) and the intrinsic estimator (see e.g Yang et al., 2008, for a review of this literature). CGLIM often specify an outcome variable as a linear function of age, cohort and period variables, and then impose some constraints on the vector of parameters. These constraints are arbitrary and are needed because the model is underidentified. The intrinsic estimator decomposes the effect of the three variables into a full rank parameter vector space b_0 and a vector which defines the linear dependence between the three variables B_0 and which is thus unrelated to the outcome variable. The full rank coefficient vector is assumed orthonormal to B_0, so that it is invariant to the selection of constraints on B_0. So in effect, this methodology also imposes constraints on the parameters of our model, indirectly by assuming orthonormality. In the economics literature, often cohorts are not accounted for or are defined as larger intervals of time than age or period (this falls under the CGLIM category of models which in this case assumes equality of cohort coefficients within a certain time interval). This works well if the changes in the experiences of different cohorts which would be relevant in happiness studies occur gradually and slowly over time, such as political and economic stability, life expectancy, social protection, and so on. However, if we define birth cohorts as ten-year intervals of birth years, we also expect their coefficients to be small and statistically insignificant because the years included in each interval are arbitrary, and so is the change from one interval to the next. This paper shows that this does not seem to be the case. Other studies have assumed linear cohort effects were zero and estimated higher order effects. Needless to say, if the linear effect is not zero, higher order effects will be biased. Other studies have used fixed effects to estimate age effects because the year of birth is a time invariant variable at the individual level (e.g Clark, 2002; Frijters & Beatton, 2008; Winkelmann & Winkelmann, 1998). However, fixed effects does not separate age from period effects, so age effects are also biased.

This paper will replicate the most common specifications of age, cohort and period effects in the economics literature. It also proposes an additional method which estimates linear, as well as nonlinear, effects of age, period and cohort, when all three variables are defined in yearly intervals. To do so, instead of imposing constraints on the parameters, it redefines the variable age. Our measure of age exploits the discreteness of the data and the fact that not all individuals are born/interviewed in the same day. As such, some individuals have had their birthday by the time of the interview while others have not. It is then possible to observe individuals belonging to the same birth year cohort with different ages purely due to exogenous reasons. This creates an exogenous source of age variation within the same birth year cohort, which breaks the linear dependence between the three variables. These linear effects, as well as nonlinear effects are thus identified with very few parametric assumptions, even at the individual level. We also try to separate the importance of other confounding factors in the age-happiness profile, such as attrition and unobserved heterogeneity. Results do differ with this method, and both attrition and unobserved heterogeneity matter in the estimation of age-happiness profiles. We also use alternative cohort effects which are not linked to the birth year of each individual, but to the year they entered school. This cohort concept has been found to be important for economic outcomes, such as educational outcomes (see e.g Pischke, 2007).

The next section describes the linear dependence problem and how linear effects of all three factors are identified. If these variables were measured continuously, and not in yearly brackets, surely the linear dependence problem would subsist. However, we argue that this redefinition of age is a better measure of age, closer to how age should be defined given that it has been discretised, and also allows for linear effects of age, year and cohort to be separately identified. Section 2.3 describes the data and how sample design of GSOEP facilitates this study. Section 2.4 estimates the age-happiness profile using alternative methods and discusses the results while Section 2.5 concludes.

2.2 Identifying the effects of age, period and *birth* cohort

We are interested in identifying the effects of age a, cohort c and period t on individual subjective well-being h. For individual i, these three factors are however linearly dependent as follows:

$$a_{it} = t - c_i, \forall t, \forall i \tag{2.1}$$

If h is well described by a general function $f\left(a, c, t\right)$ and an additively separable error term u, Eq. 2.1 implies:

$$h_{act} = f(a, c, t) + u = f(a_{ct}, t - a_{ct}, t) + u = g(a_{ct}, t) + u_c \tag{2.2}$$

Even if we would like to estimate the impact of age on happiness by conditioning the analysis on cohort and period, Eq. 2.2 shows that the initial happiness equation f can always be rewritten as a function of age and either period or cohort. To see the implications of this, let h_z represent the partial derivative of h with respect to $z, z = a, c, t$. Eqs. 2.1 and 2.2 then show that the linear effect of age on happiness h_a equals g_a. Using a chain rule, we can see that $g_a = g_t$. This is because age and time grow at the same rate, for any given cohort.

If birth cohort is omitted however, estimated effects of age will be biased in the following way:

$$E\left(\left.g_a\right| t\right) = E\left(\left.f_a\right| c, t\right) - E\left(\left.f_c\right| t, a\right), \tag{2.3}$$

From Eq. 2.3, we see that, if the birth cohort effects are positive (negative), the age effect is underestimated (overestimated).

Identifying age, cohort and period effects is an issue that arises in several different contexts. Examples include the analysis of the incidence of particular infectious diseases (e.g. Holford, 1983; Clements et al., 2005), changes in national savings ratios (e.g Deaton & Paxson, 1999), scientific productivity of researchers and vintage capital model of trucks or personal computers (e.g. Hall et al., 2005), wage structure and college premium (e.g. Welch, 1979; MaCurdy & Mroz, 1995; Card & Lemieux, 2001; B. Fitzenberger & Schnabel, 2001), human capital and early career choices (e.g. Card & Lemieux, 2000) and job satisfaction (e.g. Jurges, 2003). Different studies adopt different identification strategies. The most common type of assumption specifies each of the three variables as polynomials and restricts some of their coefficients[1]. B. Fitzenberger & Schnabel (2001); Jurges (2003); Holford (1983); Clements et al. (2005) assume the linear effects of one of the factors is zero. They then estimate higher order effects of all three factors, and their interactions. Simpler models will assume that interactions between all three factors are not important and estimate an additively separable model. This model either omits the linear effect of one of the factors, or excludes that factor from the specification altogether (Deaton & Paxson, 1999). All of these specifications have so far defined age, cohort and year in equally spaced intervals of the same length. Other authors have however proposed an additively separable model where the length of the observation periods of the three factors is no longer the same (see e.g. Card & Lemieux, 2000, 2001; Hall et al., 2005). However, Holford (1983) shows that using a model variables are defined with unequal intervals can lead to a saw-tooth profile of our parameter of interest. Finally, a less common assumption was used in e.g. Welch (1979) and Berger (1985), where cohort effects would be fully characterised by a function of cohort size. This approach relies on having a sufficient statistic for one of the factors available, which may be difficult when our variable of interest is life satisfaction. Alternatively, other authors have used an instrumental variable approach (see e.g. Heckman & Robb, 1985). They propose identifying a variable that affects the dependent variable but, in the long run, is only correlated with age, cohort or year. In the context of happiness studies, this instrumental variable also proves to be difficult to find.

[1]See MaCurdy & Mroz (1995); Hall et al. (2005) for good reviews.

In this paper, we compare different specifications of age, cohort and period effects in a linear regression model. We further propose a way of estimating all linear effects when age, cohort and period are defined in equally spaced intervals. To do so, and because data are observed on a yearly basis, age a has been redefined as completed years of life while the definition of birth cohort and period remain the same. If an individual has had his birthday by the time the data are recorded, he is $t - c$ years old. If his birthday happens later in the year, he is just $t - c - 1$ years old. Hence, as the usual measure of age in yearly longitudinal surveys, completed years of life will also be augmented by 1, but not for all individuals as soon as the calendar year changes. Depending on the exact time of the interview, individuals belonging to the same birth cohort have different completed ages in any given moment in time. This exogenous variation in the moment of the interview breaks the linear dependence between age, cohort and time, even at the individual level. This definition allows Eq. 2.1 to hold exactly for those whose birthday happens in the day of the interview. On the contrary, the usual definition of age is only close to the true relation for those who happen to be born in the first days of the year and the error increases with the lateness of the day of birth. Take individuals born in 1978 and in 1979 being observed in 1980. According to Eq. 2.1, individuals born in 1978 are all 2 years old and those born in 1979 are all 1 year old. However, individuals can have any age in the interval]0,2[if they are born in 1979 or any age between]1,3[if they are born in 1978. Our redefinition of age would assign completed years of either 0 or 1 to individuals born in 1979 and completed years 1 or 2 to individuals born in 1978.

You may argue that defining age as completed years of age is as arbitrary as defining it the usual way (which is the right way in continuous time). However, this definition breaks the linear dependence between the 3 variables, and as we will show, has smaller measurement error.

Lets define the exact age at the time of the interview as

$$\text{age}_{\text{true}} = \text{beginning current year} + s \text{ - (beginning birth year} + b \text{),}$$

where s stands for the moment of the interview and b is the moment of birth. Both variables are defined as a fraction of a given year and they are both defined in a unit interval, e.g. $s, b \in [0,1]$, where 0 means the beginning of a year and 1 the end of a year. While it is not controversial to assume $b \sim U(0,1)$, it is assumed that the moment of the interview is also equally likely in any day of the year for the sake of illustration, so that $s \sim U(0,1)$.

When age is defined as usual, i.e., as $\text{age}_{\text{usual}}$ = beginning current year - beginning birth year, the underlying error is

$$\text{error}_{\text{usual}} = b - s \in [-1,1]$$

Given the assumptions made on b and s, we know this error has zero mean and variance $\frac{1}{6}$[2].

However, when age is defined as completed years only, that is

$$\text{age}_{\text{completed}} = \begin{cases} \text{beginning current year - beginning birth year - 1} & \text{if} \quad s \leqslant b \\ \text{beginning current year - beginning birth year} & \text{if} \quad s > b \end{cases} \tag{2.4}$$

the underlying error is

$$\text{error}_{\text{completed}} = \begin{cases} b - s - 1 & \text{if} \quad s \leqslant b \\ b - s & \text{if} \quad s > b \end{cases} \in [-1,0] \tag{2.5}$$

This error has mean $-\frac{1}{2}$ and variance $\frac{1}{18}$. This paper thus proposes a biased but lower variance estimator of age[3], which breaks the linear dependence between age,

[2]The joint density of $b-s$ is $f(b-s) = 1 - |b-s|$. Hence the expected value of the error associated with the usual definition of age is $\text{E}\left(\text{error}_{\text{usual}}\right) = \int_{-1}^{1}(b-s)\left[1-|b-s|\right]d_{b-s} = 0$ and the variance is $\text{Var}(\text{error}_{\text{usual}}) = \int_{-1}^{1}(b-s)^2\left[1-|b-s|\right]d_{b-s} = \frac{1}{6}$.

[3]The expected value was computed by solving $\text{E}\left(\text{error}_{\text{completed}}\right) = \text{E}\left[(b-s)-1|\, b-s \geqslant 0\right]P(b-s \geqslant 0) + \text{E}\left[b-s|\, b-s < 0\right]P(b-s<0)$, and similarly for

period and cohort, even at the individual level. All it requires is for the moment of the interview to sometimes happen before, and other times happen after each individual's birthday[4].

2.3 Data

The German Socio-Economic Panel (GSOEP) records both the date of birth of the interviewees and the date in which interviews are held. It can happen that in a given calendar year t, individuals born in the same year and thus belonging to the same birth cohort c have different **completed years** when interviewed, depending on whether they have had their birthday by the time of the interview. Age is defined as in Eq. 2.4, and it can thus happen that two respondents from the same birth cohort have different ages in any moment in time.

As discussed in the previous section, this definition of age seems more natural given the discreteness of the data. If age is just defined as $t-c$, it is augmented by 1 just because the calendar year changed. This applies to all individuals, whether they are exactly $t-c$ years, $t-c-365$ days minus almost 6 hours old or $t-c+365$ days and almost 6 hours old. By using the definition in Eq. 2.4, age effects are not confounded with artificial "year-shifting" effects. These are identified as long as the time of the interview is purely exogenous. Individuals interviewed after and before their birthday should be identical in all except their number of completed years.

Unfortunately, only the month of birth is observed while the day of birth would provide a more accurate definition of age. In practice, age ends up being defined as $t-c-1$ if the day of the interview is prior to the 15^{th} of the month of birth and

the variance.

[4] If alternatively, we had assumed that s had a degenerate distribution at 0, so that all interviews would happen at the beginning of the interview year, the error of the usual measure of age would have expected value $\frac{1}{2}$ and the error of our proposed measure would have expected value $-\frac{1}{2}$. Both measures would have the same variance. As we move from this degenerate case to the uniform case described, our proposed measure becomes lower variance and with a higher bias than the usual measure.

$t - c$ thereafter.

Figure 2.2 shows how interviews are spread throughout the year. Interviews tend to be more concentrated in the first quarter, but they do exhibit some variation throughout the year. One source of variation is purely exogenous and stems from the fieldwork design[5]. However, there are households being contacted more than once so that their interviews tend to be carried out later in the year. If these individuals are a selected sample, who may be for instance more stressed and therefore less happy, retrials can undermine this identification strategy. For this reason, we also carry out the analysis excluding the individuals interviewed later in the year. We also run fixed effects estimation of happiness equations to account for unobserved heterogeneity.

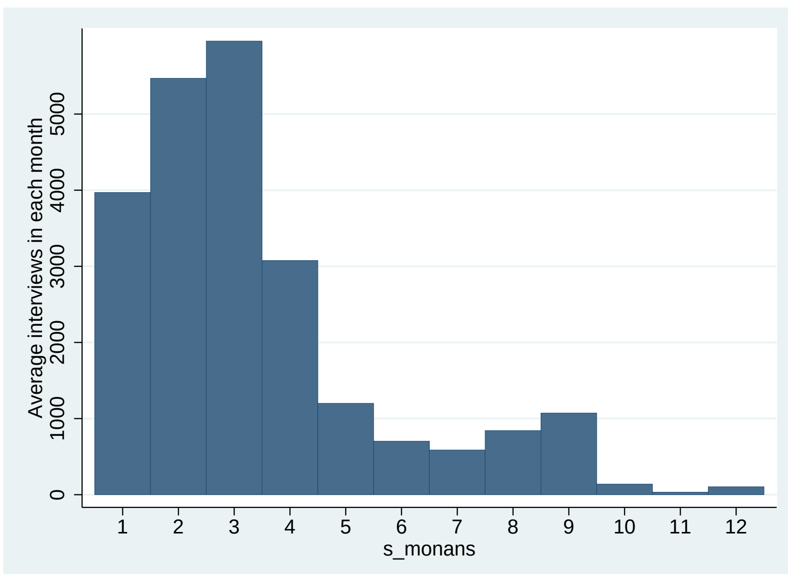


Figure 2.2: Average number of interviews conducted in each month over the 20-year period

Happiness is measured by the self-reported general satisfaction variable in the GSOEP. Interviewees are asked every year, at the end of the questionnaire, the following question:

And finally, we would like to ask you about your satisfaction with your

[5]I thank Jan Goebel from DIW Berlin for all the information regarding this issue.

> life in general. Please answer by using the following scale, in which 0 means totally unhappy, and 10 means totally happy.
>
> How happy are you at present with your life as a whole?

It is a discrete variable taking 11 integer values from 0 to 10.

Table 2.1 shows a cohort table with the sample we analyse. It represents average happiness level for individuals with a particular age in a particular year. Each row shows the evolution of the happiness mean at a given age, across time. Each column reads cross sectional values for all ages in a given period. Kermack et al. (1934) notes that lifecycle trends are observed diagonally for each cohort. With the usual age definition, each cell would correspond to a different cohort, all observed at a particular age along a row, or in a particular year, along a column, and each diagonal would represent how each cohort's average happiness evolved over time (and as they got older). With age defined as completed years, this is no longer the case. Age does not increase by 1 between interviews (years), as the evolution across any diagonal assumes for each cohort. As an illustration, we signal in bold the possible ages an individual who is 20 years old in 1986 and another who is 41 in 1985 can have in the following years. This thus shows that we can identify age, cohort and period effects, even at the individual level. A cohort is now followed along a *thick* diagonal and not a line diagonal. And this comes when age is redefined in a way which more closely matches the continuous notion of age and has lower measurement error.

Table 2.1: Average Happiness for all $(age, period)$ combinations - Cohort Table

	year of survey																		
	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
20	7.26	**7.20**	**7.30**	7.24	7.38	7.29	7.06	6.95	7.02	7.25	7.14	7.19	7.28	7.11	7.32	6.86	7.07	7.10	6.99
21	7.28	7.49	**7.22**	**7.26**	7.30	7.37	7.23	6.99	7.01	6.98	7.05	7.24	7.23	7.04	7.20	7.19	7.14	6.89	7.18
22	6.83	7.33	**7.18**	**7.18**	**7.50**	7.50	7.04	6.90	6.96	6.97	7.19	7.21	7.17	7.24	7.21	7.28	7.21	6.82	6.93
23	7.06	7.44	7.21	**7.19**	**7.14**	**7.40**	7.20	7.12	7.11	7.08	7.04	6.98	7.18	7.07	7.27	6.96	7.18	7.21	6.93
24	7.58	7.28	7.28	7.09	**7.32**	**7.36**	**7.30**	6.96	6.94	7.07	7.00	7.05	7.07	7.11	7.15	7.38	7.20	7.15	7.10
25	7.22	7.54	7.24	7.15	7.04	**7.58**	**7.13**	**7.18**	7.20	7.05	7.03	7.06	7.08	7.12	7.09	7.14	7.33	7.23	6.96
26	6.89	7.55	7.24	7.13	7.03	7.36	**6.89**	**7.11**	**7.11**	7.07	7.11	7.20	6.94	7.16	7.09	7.18	7.03	7.05	7.03
27	7.26	7.05	7.16	7.22	7.10	7.49	6.89	**6.91**	**6.93**	**7.02**	7.12	7.12	7.00	7.27	7.38	7.01	7.19	6.94	6.80
28	7.18	7.19	6.97	7.14	7.28	7.39	6.95	6.99	**7.05**	**6.97**	**7.10**	7.10	6.87	7.16	7.27	7.34	7.18	7.16	7.10
29	7.25	7.30	7.26	7.01	7.15	7.44	6.87	6.93	6.87	**6.83**	**6.89**	**6.98**	7.11	7.04	7.13	7.12	7.34	6.97	6.94
30	7.23	7.30	7.25	7.10	6.98	7.31	6.91	6.86	6.93	6.89	**6.97**	**6.99**	**6.81**	7.10	7.15	7.19	7.12	7.08	7.01
31	7.24	7.26	7.14	7.13	7.25	7.11	6.87	6.92	6.83	7.00	6.92	**7.11**	**6.84**	**7.04**	7.20	7.03	7.24	7.19	7.11
32	7.42	7.31	7.05	7.09	7.08	7.38	6.66	6.71	6.81	6.67	7.08	6.88	**6.78**	**6.87**	**7.10**	7.12	7.04	7.00	7.01
33	7.10	7.47	7.34	7.19	7.02	7.24	6.88	6.64	6.78	6.87	6.87	6.96	6.76	**6.92**	**7.05**	**7.03**	7.15	6.95	6.89
34	7.19	7.28	7.22	7.21	7.18	7.18	6.72	6.90	6.69	6.56	6.87	6.88	6.75	6.89	**7.01**	**7.11**	**7.14**	7.07	6.80
35	7.33	7.40	7.04	7.18	7.08	7.36	6.76	6.77	6.94	6.66	6.96	7.00	6.70	7.08	6.92	**6.90**	**7.05**	**6.99**	6.82
36	7.36	7.41	7.10	6.89	7.01	7.42	6.67	6.77	6.68	6.76	6.90	6.98	6.88	6.80	6.91	7.03	**7.03**	**6.82**	**6.96**

Continued on next page

Table 2.1 – continued from previous page

	year of survey																		
37	7.44	7.33	7.15	7.12	7.15	7.32	6.95	6.74	6.72	6.47	6.90	6.75	6.93	6.94	6.86	6.93	6.97	**6.83**	**6.93**
38	7.04	7.30	7.14	7.28	7.00	7.11	6.87	6.91	6.87	6.77	6.66	6.96	6.67	6.93	7.00	6.88	7.02	6.88	**6.80**
39	7.15	7.18	7.10	7.17	7.06	7.24	6.60	6.71	6.93	6.70	6.86	6.60	6.87	6.76	7.04	7.02	7.03	6.90	6.68
40	7.26	7.12	7.06	6.75	7.23	7.43	6.68	6.76	6.79	6.80	6.77	6.75	6.43	6.85	6.84	6.90	6.91	6.64	6.88
41	**7.19**	**7.20**	6.98	7.33	7.28	7.54	6.96	6.70	6.82	6.68	6.85	6.68	6.58	6.60	6.93	6.79	6.89	6.72	6.73
42	7.24	**7.27**	**7.03**	7.23	7.43	7.21	7.10	7.10	6.69	6.85	6.68	6.72	6.61	6.78	6.57	6.73	6.81	6.78	6.51
43	7.32	**7.14**	**6.96**	**6.95**	7.12	7.23	6.77	6.94	6.95	6.68	6.70	6.76	6.72	6.76	6.67	6.61	6.88	6.65	6.45
44	7.11	7.24	**7.16**	**7.13**	**6.67**	7.42	6.86	6.77	6.62	6.81	6.67	6.69	6.45	6.74	6.63	6.75	6.52	6.74	6.46
45	7.34	7.12	7.22	**6.91**	**6.77**	**7.04**	6.93	6.97	6.73	6.84	6.82	6.73	6.60	6.67	6.85	6.67	6.70	6.32	6.60
46	7.17	7.32	6.88	6.97	**7.04**	**7.19**	**6.74**	6.96	6.75	6.49	6.91	6.66	6.72	6.67	6.48	6.62	6.56	6.68	6.26
47	7.33	6.97	7.12	6.94	7.18	**7.11**	**6.77**	**6.81**	6.92	6.82	6.55	6.79	6.57	6.67	6.74	6.47	6.77	6.60	6.43
48	7.20	7.21	6.99	7.13	7.08	7.22	**6.87**	**6.64**	**6.77**	6.93	6.70	6.64	6.71	6.52	6.88	6.67	6.58	6.51	6.26
49	7.25	7.23	7.04	6.93	7.03	7.21	6.63	**6.74**	**6.54**	**6.79**	6.92	6.67	6.52	6.68	6.74	6.83	6.55	6.33	6.18
50	7.42	7.02	7.07	7.06	6.94	7.27	6.89	6.69	**6.69**	**6.46**	**6.72**	6.69	6.42	6.54	6.52	6.66	6.84	6.33	6.34
51	7.03	7.07	6.69	7.01	7.09	7.27	6.89	6.70	6.56	**6.80**	**6.48**	**6.92**	6.65	6.34	6.68	6.34	6.77	6.68	6.41
52	6.69	6.99	6.99	6.71	7.04	7.08	6.85	6.82	6.70	6.64	**6.65**	**6.45**	**6.54**	6.69	6.54	6.48	6.54	6.66	6.58
53	6.95	7.04	6.89	6.99	6.64	7.13	6.71	6.76	6.75	6.59	6.70	**6.70**	**6.23**	**6.74**	6.68	6.49	6.42	6.39	6.59
54	7.09	7.04	6.62	6.90	6.83	6.71	6.43	6.77	6.63	6.74	6.55	6.67	**6.41**	**6.41**	**6.71**	6.75	6.81	6.31	6.47
55	7.46	7.09	6.96	6.86	6.86	7.04	6.61	6.72	6.59	6.75	6.52	6.52	6.69	**6.72**	**6.46**	**6.75**	6.80	6.58	6.19
56	7.31	7.44	7.20	7.15	6.87	6.95	6.93	6.56	6.63	6.46	6.71	6.57	6.47	6.64	**6.76**	**6.41**	**6.71**	6.77	6.54

Continued on next page

Table 2.1 – continued from previous page

	year of survey																		
57	7.26	7.25	7.34	6.62	6.99	6.97	6.51	6.73	6.75	6.64	6.64	6.64	6.59	6.68	6.66	**6.69**	**6.41**	**6.55**	6.47
58	7.13	7.26	7.22	7.23	6.77	7.24	6.89	6.70	6.76	6.49	6.74	6.68	6.55	6.64	6.64	6.57	**6.97**	**6.37**	**6.48**
59	7.50	7.14	7.12	7.26	7.09	6.99	7.02	6.91	6.49	6.92	6.66	6.88	6.69	6.55	6.62	6.45	6.63	**6.68**	**6.40**
60	7.23	7.66	7.38	7.08	7.25	7.50	7.02	7.13	6.79	6.47	6.82	6.65	6.92	6.97	6.70	6.73	6.85	6.39	**6.51**

Table 2.1 also allows us to confirm some results from Fig. 2.1. It shows that, conditional on age (along each row), older cohorts are happier on average than younger cohorts, but there seem to be time effects which make some years' average happiness higher for all ages (year 1990, right after the fall of the Berlin wall, clearly shows higher average happiness for all ages, specially for younger ages). It shows that following a specific cohort (along a thick diagonal), happiness is broadly decreasing with age, but this result is not there when we look at age-happiness profiles for each year individually (each column), the same way it was not there when we pooled all years together in Fig. 2.1. This table provides additional evidence of the importance all three linear effects seem to have, and the importance of accounting for both cohort and time effects when estimating age-happiness profiles.

2.4 Estimation Results

This section shows the results of estimating happiness equations which specify age, cohort and period effects in different ways. Age is defined as in Eq. 2.4, calendar time is as usual the year of the interview, and cohort is birth cohort. We also kept the most common specifications of happiness equations for comparability of the age effects with other studies. The most common covariates used are gender, *bundesland*, nationality, marital status, number of members in the household, educational diploma, labor force status, household income and self-reported satisfaction with health. The latter is a categorical variable ranging from 0 to 10, where 10 represents full satisfaction with health and 0 complete dissatisfaction.

In order to guarantee enough observations per cell, the sample is restricted to individuals of Turkish, Balkan[6], East German or West German background, and who stay in their initial *bundesland* throughout the sample period. Those who are still in schooling, on maternity leave, have been drafted or only have a very sporadic source

[6] The countries that used to form Yugoslavia are also grouped into one category, again for sample size considerations.

of income are also excluded. Married but separated individuals are not accounted for either. Individuals are only followed after they have completed their 20 years of age and only until they reach 60 years of age. This is to prevent an over-representation of older individuals in the sample.

Table 2.2 presents the OLS estimation results. The first six columns show the results of basic specifications which do not include additional covariates. Column I shows the most common specification of happiness equations where cohorts are omitted, the age effect is modeled with a quadratic function and year dummies are included. Column II adds cohort effects by assuming constant cohort effects within 5-year intervals, as in Card & Lemieux (2001). Column III is a simplified version of B. Fitzenberger & Schnabel (2001) which models all three variable effects using cubic polynomials and assumes the cohort linear effect is zero. Column IV further includes the linear cohort effect so that we can compare and analyse the consequences of omitting the cohort linear effect. Columns V and VI use cohort and period dummies, but the former models age using a quadratic function while the latter uses age dummies. Column V is used to understand how much of the differences we observe between our estimates of the age and squared age coefficient are due to poor accounting of cohort effects while column VI tells us whether the quadratic approximation is a good one. Columns VII - XII repeat the first 6 columns but include the additional covariates. Robust standard errors are computed and errors are clustered at the individual level.

Results are striking. While the benchmark model yields the usual U-shape happiness profile with respect to age (in column I, the inflexion point occurs around age 76, outside our age range; in column VII, it occurs in the late 30s), no other specification which accounts for cohort and time in some way replicate this result. When using 5-year cohort intervals or polynominal functions, the age-happiness profile estimated is inverted-U, with maximum happiness around the 20s or 30s (some profiles are actually decreasing, given that the inflexion point is estimated to be at an age outside our sample range). When cohort and time dummies are included, the predicted profiles are increasing. Exception is the specification with added covariates and age dummies, when the age-dummy variables stop being statistically significant. So while the quadratic specification seems to be forcing a hump which the age dummies do not confirm, there is not enough variation in the data to identify the effects of age, cohort and time dummies in the last column (only the 1991 dummy variable is significant from the results shown). There are additional points about time and cohort effects worth mentioning. When we add cohort 5-year interval dummies, cohort effects do not seem to vary much, which is to be expected given the arbitrariness of the cutoff points. However, cohort polynomials or dummies do point to statistically significant positive cohort effects in the basic specification, suggesting individuals born later are on average happiest, even if these cohort effects are captured by socio-demographic characteristics whose changes correlate with differences across cohorts. Time effects are estimated to be negative, even if their significance wanes in more saturated and more flexible specifications. But it is surprising to see cohort effects being estimated as positive, suggesting that those born later, the younger cohorts, are happier. Visual inspection of Fig. 2.1 and of Table 2.1 would suggest otherwise. But neither Fig. 2.1 nor Table 2.1 represent the effects of each of the three variables independently from the others.

The estimates of the additional covariates do not yield surprising results[7]. Household net income has a very significant albeit small impact on happiness. The divorced individuals fare worst and the widowed are worse off than single individuals, even though age and satisfaction with health are in the equation. Households with 4 members or more are doing poorly, even after conditioning on income. The unemployed are the least happy group while the Full-time workers and the retired individuals are the happiest. Men are significantly less happy than women. Similar to other studies, educational differences are not statistically significant. There are also important regional and nationality differences. Health is the most important factor in explaining happiness.

[7] These are available upon request.

All in all, estimating a happiness equation with age redefined and without conditioning on birth year still yields a robust U-shape profile. Results indicate however that the age coefficient estimates from previous work are in fact a combination of positive cohort effects and negative age effects, specially for older ages. Looking at the standard errors of the age coefficients, one further sees that the true explanatory power of age is very reduced, once year of birth is adequately accounted for in the analysis. This can also be due to lack of variation in the moment of the interview, because all three variables lose explanatory power when more saturated models are used.

Table 2.2: OLS estimates of a happiness equation, conditioning on age, birth cohort and calendar time

	I	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**
Age	-0.0304***	0.0192**	0.1590***	0.2381***	0.1244***		-0.0447***	0.0013	0.1659***	0.1465***	0.0024	
	(0.0050)	(0.0090)	(0.0260)	(0.0360)	(0.0270)		(0.0050)	(0.0080)	(0.0240)	(0.0310)	(0.0200)	
Age^2	0.0002***	-0.0003***	-0.0030***	-0.0034***	-0.0003***		0.0006***	0.0002**	-0.0039***	-0.0037***	0.0001*	
	(0.0000)	(0.0000)	(0.0010)	(0.0010)	(0.0000)		(0.0000)	(0.0000)	(0.0010)	(0.0010)	(0.0000)	
Age^3			0.0000***	0.0000***					0.0000***	0.0000***		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Age 21						0.1408***						0.0514
						(0.0460)						(0.0430)
Age 31						1.2882***						0.2267
						(0.2890)						(0.2200)
Age 41						2.2815***						0.3117
						(0.5490)						(0.4150)
Age 51						3.1213***						0.2975
						(0.8120)						(0.6130)
Age 60						4.1524***						0.6417
						(1.0470)						(0.7890)
[1929, 1934[		-0.1374						-0.0001				
		(0.0950)						(0.0780)				
[1939, 1944[		-0.204						-0.1581				
		(0.1340)						(0.1040)				
[1949, 1954[		-0.1718						-0.1608				
		(0.1990)						(0.1530)				
[1959, 1964[		-0.0423						-0.0529				

Continued on next page

Table 2.2 – continued from previous page

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
		(0.2690)						(0.2040)				
[1969, 1974[		0.216						0.1663				
		(0.3370)						(0.2550)				
[1979, 1984[		0.2585						0.314				
		(0.4060)						(0.3070)				
Cohort				0.0765***						-0.0187		
				(0.0280)						(0.0210)		
Cohort^2			0.0010***	0.0005					0	0.0001		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Cohort^3			0	0					0.0000**	0.0000*		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Born 1925					0.2634	0.3225					-0.1089	-0.0934
					(0.2660)	(0.2660)					(0.2530)	(0.2540)
Born 1935					0.7856**	0.9843***					-0.4035	-0.2875
					(0.3670)	(0.3670)					(0.3110)	(0.3110)
Born 1945					1.8715***	2.0816***					-0.3463	-0.2157
					(0.5970)	(0.5970)					(0.4700)	(0.4690)
Born 1955					2.8162***	2.9727***					-0.4104	-0.3439
					(0.8430)	(0.8430)					(0.6490)	(0.6490)
Born 1965					4.0657***	4.2213***					-0.2338	-0.1691
					(1.0930)	(1.0920)					(0.8360)	(0.8350)
Born 1975					5.4411***	5.6225***					0.121	0.2141
					(1.3510)	(1.3510)					(1.0270)	(1.0260)
Born 1983					6.6581***	6.8727***					0.4166	0.5395
					(1.5770)	(1.5750)					(1.2600)	(1.2590)
Year 1986	-0.0129	0.0065			0.4118***	0.4138***	-0.0329	-0.0223			-0.031	-0.0299
	(0.0260)	(0.0400)			(0.1090)	(0.1090)	(0.0260)	(0.0350)			(0.0840)	(0.0840)

Continued on next page

Table 2.2 – continued from previous page

	I	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**
Year 1991	-0.4011***	-0.4103***			-0.5054***	-0.5052***	-0.1821***	-0.1894***			-0.1829***	-0.1829***
	(0.0240)	(0.0250)			(0.0360)	(0.0360)	(0.0230)	(0.0240)			(0.0310)	(0.0310)
Year 1996	-0.4105***	-0.4731***			-1.0721***	-1.0729***	-0.1524***	-0.1934***			-0.1806	-0.1838
	(0.0250)	(0.0500)			(0.1580)	(0.1580)	(0.0240)	(0.0410)			(0.1200)	(0.1200)
Year 2001	-0.3321***	-0.4576***			-1.5587***	-1.5591***	-0.1273***	-0.2217***			-0.2028	-0.2037
	(0.0270)	(0.0850)			(0.2880)	(0.2880)	(0.0260)	(0.0650)			(0.2170)	(0.2170)
Year			-0.1504***	-0.2184***					-0.0868***	-0.0701***		
			(0.0150)	(0.0300)					(0.0140)	(0.0240)		
$Year^2$			0.0062***	0.0066***					0.0072***	0.0071***		
			(0.0010)	(0.0010)					(0.0010)	(0.0010)		
$Year^3$			-0.0001***	-0.0001***					-0.0002***	-0.0002***		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Constant	8.0714***	7.0747***	4.0644***	0.0652	0.5441	2.6856**	4.9267***	4.0280***	1.5889***	2.5630**	4.2898***	4.2309***
	(0.1020)	(0.4730)	(0.5480)	(1.5680)	(1.7150)	(1.1980)	(0.1310)	(0.3740)	(0.4640)	(1.1960)	(1.3020)	(0.9130)
R^2	0.014	0.016	0.014	0.014	0.018	0.018	0.292	0.293	0.293	0.293	0.294	0.295
Covariates	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Profile	Decreasing	Inverted U	Inverted U	Inverted U	Increasing	Increasing	U-shaped	Increasing	Inverted U	Inverted U	Increasing	Unrelated
Inflexion Point		36.4	26.9	34.9			34.5		21.5	19.6		

Significance levels : * 10% ** 5% *** 1% Standard errors in parentheses

Additional covariates are gender, *bundesland*, nationality, marital status, educational diploma, labor force status, household income and self reported satisfaction with health and number of members in the household.

Omitted categories: 20 year olds, year 1984, cohort born between [1924, 1929[, cohort born in 1924.

2.4.1 Attrition, Unobserved Heterogeneity and Ordinal nature of happiness

The previous estimation results are subject to a number of criticisms. First of all and as already discussed, the exogeneity of the moment of the interview only holds if all the interviewees answered the first time they are contacted or if the reasons why they might not have replied in the first attempts are uncorrelated with happiness, conditional on all covariates[8]. Interviews being carried out later in the year might be contaminated with those individuals who are less available and with a higher valuation for time. In fact, Frijters & Beatton (2008) showed that there seems to be selective attrition and the average happiness of those who stay in the panel is lower than the overall average. If we think that those who need to be contacted again are also more likely to attrite in the future, we should worry. We reestimate the happiness equations for those that are interviewed only in the first months of the year to avoid including interviews where respondents had to be contacted more than once. We also look at those who stay in the panel for the whole 20 waves and also for those who answer the first and the last questionnaires. Finally, we analyse how results change when we account for unobserved heterogeneity and/or the ordinal nature of the happiness variable by running fixed effects, ordered probit and ordered fixed effects logit estimations[9]. Results are in Appendix E.

Late interviews

The regressions are repeated for only the first months of the year. This aims to withdraw from the sample those individuals who have to be contacted more than once because their interviews tend to be concentrated later in the year. Tables 9, 10 and 11 show the estimated age-happiness profiles when only the first three, four and

[8]The number of attempts made for each interviewee is actually a piece of information which should be made public.

[9]We thank Paul Frijters and Ada Ferrer-i-Carbonell for useful discussions about their method explained in Ferrer-i-Carbonell & Frijters (2004) and for having made their Stata code available. All errors are my own.

six months respectively are used for estimation. In short, all results remain qualitatively the same, which indicates that results do not seem to vary according to the number of attempts made before the interviews take place. In the basic specification, an inverted-U shaped age-happiness profile seems to emerge, and the inflexion point remains at very young ages. However, in the full more flexible specifications, age, cohort and time effects remain statistically insignificant.

Stayers

The happiness equation is also estimated with a balanced sample to account for a possible selection bias. First only those individuals who answer all of the questionnaires are included and results are presented in Table 12. Only 2273 out of 33852 individuals satisfy this condition and so, the exercise is repeated with all the interviewees who answered the first and the last questionnaire. This more than doubles the number of individuals. Table 13 shows these results.

For both samples, the benchmark model continues to present a statistically significant U-shaped age-happiness profile. Most models where age is a quadratic function continue to exhibit an inverted U-shaped profile, except for the model with a complete set of cohort and period dummies, which do not show a statistically significant relation between happiness and age. The inflexion points remain at very young ages, so age-happiness profiles seem to be decreasing. Using a balanced panel does not seem to change the nature of the results, contrary to Frijters & Beatton (2008).

Alternative Estimation Methods: accounting for selection and the ordinal nature of the happiness variable

Results have also shown that age and time effects are also important and do not vary collinearly. Given the advantages of a sampling design which exploits age variation also at the individual level, this section shows the results from fixed effects estimation, which accounts both for cohort effects, unobserved heterogeneity and attrition bias. It also estimates our happiness equation using ordered probit to account for the ordinal nature of the happiness variable. It further estimates an ordered fixed-effects logit equation to simultaneously account for both issues. Table 14 shows the results.

Within Groups estimation is carried out. With age defined as in Eq. 2.4, the age and calendar time no longer grow at the same rate at the individual level, which makes it possible to estimate age effects separably accounting for period effects, while cohort effects are removed with the fixed effects. Results change dramatically. The quadratic specification of age, including the additional regressors, now suggests a clear decreasing age-happiness profile (inflexion point of a now decreasing profile is very high). This result is confirmed by the specifications which use age dummies. At the same time, year effects are also statistically significant for most years. The negative profiles had already been found in Clark (2002), but he could not separate year from age effects. These results are confirmed by the ordered fixed effects logit results. Results do seem to be driven by fixed effects, since our ordered probit estimation results do not reproduce them. In fact, with the probit estimates, age does not seem to matter for happiness and cohort effects seem to be negative, which runs counter to our OLS results from previous sections. So it seems that accounting for fixed effects reinforces previous results suggesting a decreasing profile, or at best, an inverted U-shaped profile with a maximum average happiness at relatively young ages. This suggests that apart from important cohort effects, the estimation of happiness equations needs to take selective attrition seriously, because results have more clearly shown a decreasing profile than an inverted-U shape.

2.4.2 Using school year cohorts

Our identification strategy does rely on enough variation in the time of the interview so that we observe enough variation in age within the same birth cohort, for any period. The more flexible specifications render all effects not significant, which may be an indication that the data does not have enough variation. Results have shown that polynomials of cohort show strong linear effects, which is not captured by cohort dummies which assume cohort homogeneity within 5 or 10 year intervals. So this paper invites a sampling design which varies the time of the interview throughout the periods. This section however proposes an alternative way of measuring cohort effects which does not rely as much on the sampling design and which is likely to produce more significant results. It will use the cohort defined as the group of individuals who starts their primary school in the same year. Evidence suggests that there are significant differences observed in consecutive school cohorts defined this way (e.g Pischke, 2007). Because this measure of cohort is not linearly dependent on age and year, using this measure also allows us to use finer data on age and year. This section shows how results change with this measure of cohort, using the full specification and dummy variables for age, cohort and period. We will then use finer data on age and time and change their annual frequency to monthly frequency. Due to the small number of interviews taking place in the second half of each year, all observations from the second semester have been grouped together. Results are in Table 15. Using a schooling cohort measure, we confirm that the age-happiness profile is decreasing. The impact of age and period on happiness become stronger with this measure of cohort, but the cohort effects themselves are quite negligible. When we use higher frequency for age and period, cohort effects are more precise and seem to be positive, but age and period effects disappear. While age and time should be thought of as continuous, monthly frequency is too high to show any significant effects on happiness, despite accounting for potential fluctuations in happiness which could repeat themselves annually[10]. This specification had virtually no impact on the R^2 of the model.

[10] Given the high frequency of age and period, results are not shown in the appendix, but available upon request.

2.5 Conclusion

This paper revisits the age-happiness profile and focusses specifically on how the specification of cohort effects impacts on the results. Accounting for age, cohort and period effects is always a challenge due to their linear dependence. We discuss the relative merits of alternative specifications and compare their results. We also propose an alternative definition of age which allows for individuals from the same birth year to be observed in a given year with different ages. When data are observed on a yearly basis, and relying on the fact that not all individuals are born nor interviewed on the same day, we can observe individuals born in the same year with two different ages in a particular moment. Defining cohort and period the usual way, but redefining age as completed years of age at the time of the interview breaks the linear dependence between the three factors. OLS results suggest that average happiness has an inverted U-shape, with a maximum happiness at young ages. When fixed effects are included, the estimated age-happiness profile is decreasing. Using an alternative measure of cohort which groups together individuals who have gone through their early schooling together reinforces these results. Together these results show that the U-shaped profile often found when cohorts are omitted is no longer observed. This implies that cohort effects, even if not always significant, can have a substantial impact on the variable of interest and omitting them or inadequately accounting for them can render conclusions invalid.

The key element to implement this procedure is having enough variation in the month of the interview and the recording of individual birthday, preferable the day of birth which is not however available in this dataset. As long as adequate accounts of time have been made, spreading interviews throughout the year allows the econometrician to observe two individuals that are exactly the same in everything except in their number of completed years. Further, interviewing each individual in different moments of the year further allows the same individual being observed in two consecutive years with the same age or a 2-year difference in age. Moreover, recording the number of attempts made, before succeeding in contacting the interviewee, would help identifying the group of people most likely to bias the results.

Skepticals may wonder that whichever way we find to account for age, period and cohort is always arbitrary because in continuous time, these three variables are still linearly dependent and only non-testable assumptions can allow us to estimate their impact. The point is that we are redefining age in a way which is not worse than the usual definition but has the benefit of allowing us to analyse the linear effects of three fundamental variables. We find that this is a route worth exploring and interview design should allow this to happen.

Chapter 3

Costs of domestic violence: a life satisfaction approach

Abstract

This paper discusses and estimates the costs of domestic violence using a life satisfaction approach. It draws on a British cross sectional dataset which includes individual self-reported life satisfaction, household income and experienced domestic violence, and estimates the costs of domestic violence as the compensating variation of domestic violence resulting from estimating a life satisfaction regression equation. Some attempts to account for self-selection into abusive relationships, and for the endogeneity of household income are discussed and implemented. Results suggest domestic violence is costed very highly by its victims, with estimates ranging from as little as £1000 up to over £50000. In the aggregate, compensation for domestic violence accounts for a significant percentage of total GDP. Hence this paper contributes to the literature on valuing non-marketable goods and discusses the usefulness of a life satisfaction approach when estimating the costs of domestic violence. It claims that despite its shortcomings, a life satisfaction approach allows for a valuation of the costs of domestic violence and provides answers often other valuation methods fail to.

JEL classification: D1, I3, J12, O15

Keywords: individual costs of domestic violence, compensating variation, life satisfaction approach.

3.1 Introduction

One of the major challenges of public policy is to value non-marketed goods and services, without which governments cannot make informed choices about how to allocate public spending. The absence of a price determined by a relevant market means that valuation methods used to estimate the costs of non-marketable goods and services are fraught with difficulty. This paper attempts to estimate the costs of one such non-marketed good, domestic violence, whose effects on the victims' integrity, economic outcomes, and mental health are overbearing. It will do so by estimating the compensating variation of domestic violence resulting from estimating a life satisfaction regression equation.

There are three main valuation methods of non-marketable goods at the individual level, revealed preference methods, hedonic regression, and stated preference methods. Revealed preference methods have been used, for instance, in Rao et al. (2003), who estimates the cost of safe sex as the price penalty prostitutes incur for using condoms with their clients. This method relies on there being a natural experiment which identifies a counterfactual group of people not exposed to the same treatment, which may not always exist. Gibbons & Machin (2008) uses a hedonic regression analysis to estimate the value of public services and school quality. This method relies on there being a marketable good, such as housing, whose price changes systematically with the quality of the non-marketed good, in this case both public services and school quality taken together. As long as house prices are in equilibrium, as long as houses only differ to the extent that they are located in areas with differing exposures to the non-marketed good, and as long as the data are good enough and allow for individual self-selection to be accounted for, house prices will reflect the non-marketed good's value. The third valuation method often used in valuation is somewhat different. Instead of relying on observed data to reveal information about the non-marketed good, it asks respondents directly about how they value it. Stated preference methods have been applied to assess the value of different types of crime. Atkinson et al. (2005) has estimated that different types of crime can cost each victim up to £36000 in the UK. However, asking individuals direct questions about their valuation of a specific good invites strategic responses, and can give rise to unreflective or idiosyncratic answers framed by the particular context of the question. Moreover, there is evidence that average individual self reported willingness to pay does not often have the same magnitude average individual willingness to accept (see e.g. Knetsch, 2000).

In the context of domestic violence, its valuation is as important as it is challenging. Natural experiments which would randomly allocate individuals to different incidence levels of domestic violence may be rare if at all possible, and randomised trials which could fabricate such variation are rare (an exception is Hidrobo & Fernald, 2013). Hedonic regressions rely on there being a marketed good whose price changes with domestic violence, which even if existing, would then require strong assumptions in terms of market equilibrium, and large demands on data quality to isolate the price variation attributed to domestic violence only. Stated preference methods, despite its limitations, have been used to estimate costs of crime. In England and Wales, Walby (2004) has estimated the costs of domestic violence at the national level, following a methodology proposed in Brand & Price (2001). They combine accounting techniques and stated preference methods to estimate different types of costs. Economic costs were estimated mostly by modeling and costing the relations crime has with marketed activities, or with outcomes such as industry turnover and absenteeism, while emotional costs were estimated using stated preference methods drawing on data from the British Crime Survey (BCS). Brand & Price (2001) estimate that the total cost of crime in England and Wales was 60 billion sterling in 2000. Walby (2004) finds that the costs of domestic violence alone were 20.06 billion sterling in 2006/7, out of which 13.88 billion were human and emotional costs.

This paper offers an alternative valuation method of domestic violence. Relying on individual data on self-reported life satisfaction, household income and experienced domestic violence, it estimates a life satisfaction regression equation dependent on income and domestic violence. Individual costs of domestic violence are derived as its estimated marginal rate of substitution with respect to household income. While not suffering from most of the limitations of more conventional valuation methods, it has limitations of its own. This paper assumes self-reported life satisfaction is a good indicator of utility and is the ultimate variable to maximise. Sen (1990) argues that self assessments of life in general include adaptation and levels of resignation which invalidate the use of this variable. Others argue that, because self-reported satisfaction "is a global retrospective judgement, which in most cases is constructed only when asked and is determined in part by the respondent's own mood and memory, and by the immediate context" (Kahneman & Krueger, 2006), it is inadequate in assessing individual overall well-being, and in comparing responses across individuals. Despite these drawbacks, there is vast research from Psychology validating life satisfaction data against more objective measures of emotional state (see e.g. Clark et al., 2006). There is also mounting evidence showing that the relation between life satisfaction and several important socio-demographic and economic factors is stable across different studies (see e.g Frey & Stutzer, 2002); and that major events in a lifetime, such as divorce, job loss, or bereavement, often have permanent effects on one's life satisfaction (see e.g Lucas et al. (2003) for a discussion of the impact of transitions in marital status); and that the importance of different domains of life, such as health, intimacy, or material well-being, is also relatively stable (see e.g Cummins, 1996). This paper assumes that it is reasonable to make these assumptions and explore the benefits of engaging with life satisfaction data in furthering our understanding of the weight domestic violence has on well-being.

Estimating consistent estimates of the effect of domestic violence and income on life satisfaction has additional caveats. To begin with, studies have often found a weak relation between life satisfaction and income (an example is the seminal work from Easterlin, 1974, which shows this weak correlation when looking across different countries; but similar evidence has been found when looking at time series data for a particular country, and for longitudinal data also). Individuals seem to adjust to changes in income very quickly and often completely, specially as a result of positive changes to income (Clark et al., 2006). If the degree of adaptation and social comparison effects are this strong, then there would not be a monotonic relation between income and utility, and the compensating variation of domestic violence would be ill-defined. More recent studies have however shown that, when the endogeneity of income is accounted for, its effect on life satisfaction in longitudinal studies is large and more significant (see e.g Powdthavee, 2009). Given the limits of the data, this paper will therefore attempt to account for the endogeneity of income and argue that adaptation may partly be accounted for by the inclusion of personality variables. We will include an imputed potential wage and local crime rates based on postcode information which can partly account for social comparisons, even if no systematic analysis of social comparisons and reference groups is being made.

It is also very likely that there is endogenous selection of exposure to domestic violence. Pollak (2002) develops an intergenerational model of domestic violence which explains the perpetuation of violence in homes where victims have been exposed to and therefore tolerate violence more. Part of the issue has to do with people conforming to their circumstances and there being personalities which tolerate abusive behaviour more than others (e.g Lundberg, 2010, shows how more agreeable people tend to divorce less). We assume that the personality variables will significantly reduce the impact of this source of bias. Local crime rates also proxy for exposure to crime and erosion of social norms.

The next section briefly summarises the methodology. Section 3.3 describes the data, alerting to the challenges that the data available add to this exercise. Section 3.4 presents and discusses the estimation of the marginal utility of income and violence, while section 3.5 presents the estimates for the individual and aggregate costs of domestic violence. Section 3.6 concludes.

3.2 Life Satisfaction Approach

Recent years have seen an increased interest in the economic consequences of domestic violence and on its social and private costs. Bowlus & Seitz (2006) shows that abused women are more likely to divorce and less likely to be employed. With a dynamic model, it also suggests that once violence has taken place, increasing women's employment may in fact worsen the incidence of domestic violence. Morrison & Biehl (1999), in turn, shows how children that have been exposed to domestic violence tend to underperform at school, making the economic effects of domestic violence intergenerational and long lasting. Pollak (2002) went one step further and modeled the propensity to tolerating and perpetrating violence as a function of previous exposure to violence. He concluded that violence does tend to stay in families previously exposed to it. Tauchen et al. (1991), Farmer & Tiefenthaler (1997) and Aizer (2007) find that domestic violence is more likely to occur the lower the economic opportunities of the victims. More recently, Hidrobo & Fernald (2013) shows that cash transfers received by women in Ecuador decrease domestic violence for higher education groups, but for lower education groups, it can actually increase if the woman's education is at least as high as the man's. Given that domestic violence is one of the most costly types of crime and one of the main sources of crime suffered by women in the absence of armed conflict, this paper provides an estimate of the total costs of domestic violence for the victims using a methodology that has not been used so far.

Our approach assumes self-reported life satisfaction is a good proxy for utility and estimates a utility function U which depends positively on household income y and negatively on domestic violence DV. The compensating variation for domestic violence CV can be obtained by equating utility in a non-violent state 0 with utility in a violent state 1. $U^0\left(y^0, DV^0\right) = U^1\left(y^0 + CV, DV^1\right), CV \geq 0$. With a separable happiness equation as follows

$$E\left(U_i \mid DV_i, y_i, X_i\right) = \alpha_0 + \alpha_1 DV_i + f\left(y_i\right) + \alpha^{'} X_i + \varepsilon_i \tag{3.1}$$

where X represents all additional covariates, CV will solve the equation

$$E\left(U_i \mid DV_i = 0, f\left(y_i^0\right), X_i\right) = E\left(U_i \mid DV_i = 1, f\left(y_i^0 + CV\right), X_i\right) \tag{3.2}$$

The most common functional forms used in the literature for the income function are the linear and the logarithmic forms. Both impose relatively strong assumptions on the relation between the compensating variation and the level of income. While the linear form assumes all victims of domestic violence would require on average the same compensating variation to neutralise the effects of violence, regardless of their household income, the logarithmic form assumes an increasing relation between CV and household income[1]. This paper explores two alternative specifications, the Box-Cox transformation and a quadratic function of income. The former nests the linear and the logarithmic forms and can test whether they are good approximations.

While never used to calculate the costs of domestic violence, this approach underlies the estimation of the tradeoff between unemployment and inflation discussed in Tella et al. (2001). Other applications of this approach now include a valuation of droughts and floods (Carroll, Frijters & Shields, Carroll et al.), informal care (van den Berg & i Carbonell, 2007), death of a loved one (Deaton et al., 2009; Oswald & Powdthavee, 2007), urban renewal (Dolan & Metcalfe, 2008), air quality (Luechinger, 2009; van Praag & Baarsma, 2001) and terrorism (Frey et al., 2004).

3.3 Data

The main dataset of this paper is discussed in Anand et al. (2009). It was designed to demonstrate the notion that capabilities can be measured, taking a leap towards

[1]If $f\left(y_i\right)$ is linear with parameter α_2 in y_i, CV equals $CV = -\frac{\alpha_1}{\alpha_2}$. If $f\left(y_i\right) = \alpha_2 \log\left(y_i\right)$, CV equals $CV = y_i^0 \left(\exp^{-\frac{\alpha_1}{\alpha_2}} - 1\right)$.

operationalising Sen (1993)'s capabilities approach. The design of the questionnaire relied on Nussbaum (2000)'s list of capabilities, and contains a set of 65 capability indicators together with a rich array of socio-demographic and economic variables. The survey instrument was delivered in 2005, between the 17^{th} and the 22^{nd} February, to a subsample of approximately 1048 individuals of the UK YouGov database. It was administered online and it is anonymous. This is, despite its modest size, one of the few datasets which includes information on experienced violence that does not come from a self-selected sample of reported victims.

The data set contains two main variables on experienced domestic violence. The wording of the most robust variable is as follows.

> Have you ever been a victim of domestic violence (yes=1/no=0) (Domestic Violence ever)

Victims of domestic violence often do not report incidents either to conform with social norms, for fear of consequences (Moreno et al., 2005), or because they may have altruistic preferences for the perpetrators and may not want to expose them. Because this survey is anonymous and administered online, it is less likely that respondents will misreport their domestic violence experiences than it is in other existing data sets. Jarvinen et al. (2008) claims 1 in 4 women will experience an act of domestic violence in their lifetime. Our data suggest a similar incidence of domestic violence for women, and a not so negligible incidence for men. Out of the initial 1048 respondents, 15 people did not provide an answer to this question. From the 1033 respondents, 22.8% of women report having been a victim of domestic violence and this percentage is almost 10% for men.

This paper also compared the incidence rates of this data set with the incidence rates from the self-completed British Crime Survey intimate personal violence (IPV) module. The IPV module asks two different questions about experienced violence. These questions are asked to all individuals in the sample, men and women, aged between 16 and 59 years old. The questions list the types of offenses, from verbal abuse to sexual abuse, the victim may have suffered, and respondents have to select yes or no to each item individually. It asks about experiences in the last 12 months prior to the interview, and about experiences since the age of 16. The question which mirrors more closely our first measure of experienced violence is the latter. In 2009/2010, 15.8% of men reported having been victims of domestic violence and this number grew to 17% a year later. For women, the percentage of victims varied from 29.4% and 29.9% in this period (Chaplin et al., 2011). These percentages are higher than the percentages of our dataset, but this may be due to differences in the structure of the questions. The questions in the IPV BCS module, when changed from a list of offenses to a yes/no question on each type of offense, seems to have increased the percentage of respondents answering affirmatively (Hall & Smith, 2011). The fact the question in the data set used in this paper is an even coarser question may justify a slightly lower incidence.

This first measure of experienced violence is a bit unclear for the purposes of our paper because we do not know how long ago or how frequent and severe the incidents were, nor do we know whether they are still happening. The data set also includes a measure of vulnerability to domestic violence, which asks respondents to provide a number from 1 to 7 to represent how vulnerable they feel to future violence in their home (7 being the most vulnerable).

The actual wording is as follows.

> Please indicate how vulnerable you feel to domestic violence in the future - using a scale of 1 to 7 where 1 means "not at all vulnerable" and 7 means "very vulnerable"?

Table 3.1 shows how respondents who report having been victims of domestic violence or not answer the question about vulnerability to domestic violence. Everyone answered this question. Out of the 174 respondents report having been victims of domestic violence, only 78 report even the mildest vulnerability to future domestic violence (an answer larger than 1), and less than 10% reports extreme vulnerability (an answer of at least 6). From the 859 respondents who report no past incidents with domestic violence, only 52 report a number higher than 2. So while vulnerability to domestic violence is a more informative measure of how pervasive this experience is at the time of the interview, the number of people reporting any vulnerability is rather low. What is more, it makes comparisons between answers more difficult as there may be more scope for different interpretations of the notion of vulnerability. This paper uses both measures of experienced violence.

Table 3.1: How vulnerable to current and future domestic violence is the sample?

Vulnerability at home	Not at al	2	3	4	5	6	Very vulnerable	Total
Never victims of DV	711	96	22	15	10	3	2	859
Victims of DV	96	28	12	12	10	11	5	174
Total	807	124	34	27	20	14	7	1,033

For estimation purposes, the vulnerability variable will be collapsed into a binary variable, which will take the value 1 for all individuals who report vulnerability to domestic violence at least as high as 4, and 0 otherwise. Using this variable, the percentage of people who are currently subject to domestic violence is 7.06%, which represents a 4.90% for men and a 8.81% for women. Table 3.2 shows the percentage of IPV respondents who said they had been victims of domestic abuse in the last 12 months prior to the survey. This measure of violence is likely to compare more closely to our measure of vulnerability because on average respondents who have had recent incidents should also report higher vulnerability. The table shows that the percentage of women reporting recent experiences varies between 6.4% and 8.2% in the period 2004-2011, while for men, these percentages vary between 3.6% and 5.8%. Both IPV and our data set therefore produce similar magnitudes of the incidence of domestic violence more likely to be included in respondent's assessment of current life satisfaction.

Table 3.2: Percentage of victims of domestic violence in last 12 months

	2004/05	2005/06	2006/07	2007/08	2008/09	2009/10	2010/11
Men	5.6	5.5	5.8	5.2	4.0	3.6	4.0
Women	8.1	8.2	7.9	6.9	6.4	6.6	6.4

Source: (Chaplin et al., 2011)

It is well known that what is meant by domestic violence varies across people of different educational and social background, income levels, but mainly, of different sex. While domestic violence for women often entails physical abuse, domestic violence suffered by men is almost always of a verbal and emotional nature. Comparing answers between men and women is therefore problematic. However, the small percentage of men who report experienced domestic violence, makes their separate analysis more unreliable. This paper presents the results separately just for women, for men, and for the whole sample.

The self-reported measure of life satisfaction is the answer to the question

> How satisfied or dissatisfied are you with your life as a whole?

The question is clearly aiming at an overall appreciation of one's life, so it can be argued that it is a good measure of utility. What is not so clear is what is meant by life as a whole. It is not clear if it invites an analysis of *current* life as a whole, or life as a whole *until now*. This ambiguity not only adds to measurement error because different respondents may have read the question differently, but what is perhaps more worrying, is that it makes the analysis of our coefficients, and the estimation of the costs of domestic violence much less clear. We have assumed that the answers represent an integral of how people perceive their lives *until now*, so that we estimate the CV as the change in this measure caused by the flow of violence. This question is asked both at the beginning and at the end of the survey. Several studies (e.g. Pudney, 2010) show how values of satisfaction vary significantly with the location of the question in the questionnaire. This paper uses the second measure on the grounds that it should be less subject to idiosyncracies and current mood because it comes after the respondents had to reflect on several relevant areas of their lives. This will be our measure of utility.

The income variable included in the data set is gross household income, a more natural measure of income, specially for women living in traditional households. The questionnaire includes the following question:

> Gross household income is the combined money income of all those earners in a household including wages, salaries, or rents and BEFORE tax and contributions to national insurance are deducted. What is your gross household income?
>
> - 0 - nothing
> - £1 to £9,999 per year (£1 to £199 per week app)
> - £10,000 to £19,999 per year (£200 to £389 per week app)
> - £20,000 to £29,999 per year (£390 to £574 per week app)
> - £30,000 to £39,999 per year (£575 to £774 per week app)
> - £40,000 a year or more (£775 a week or more)
> - Prefer not to answer
> - Don't know

Over 4% of respondents said they did not know their household income and over 10% chose not to answer, so the sample with non-missing household income reduces to 883 respondents. While income data provided as an interval makes it more likely respondents will answer truthfully, this study needs a continuous measure of income. What is more, previous studies have shown that not accounting for the endogeneity of income in life satisfaction regressions tends to underestimate the effect of income, and is claimed to be the reason why the estimated relation between income and life satisfaction is often not statistically significant. In this paper, this underestimation would lead to an overestimation of the costs of domestic violence. For these two main reasons, the estimation of the marginal utility of income is the major weakness of this paper. We use two continuous measures of gross household income based on the gross household income data from the British Household Panel Survey (BHPS). We replace each income band value, from 1 to 6, by the BHPS average income in each interval. Layard et al. (2007) uses the midpoint of each income band instead but, given the positive skewness of the household income distribution, the mean imputed from a comparable data set can be argued to be a better starting point. The negative correlation between income and domestic violence does mean imputation exercises based on a data set without information on domestic violence is likely to overestimate the imputed value of victims of violence. Nevertheless, this being an unconditional average, we argue this may not be a major problem. This measure of household income will only have 6 distinct values, which is more worrying. For this reason, we have also used an alternative measure of gross household income which is an imputed value from the BHPS, after matching individuals between the BHPS and the dataset in this paper based on observable characteristics. The survey used was designed using very similar questions to the BHPS, so not only are the two measures of gross household income comparable, but so are most of the relevant characteristics.

The lack of data on experienced violence from the BHPS means that income of victims is likely to be overestimated. In our imputation exercise, we tried to estimate a fixed effects regression equation, by using BHPS waves from 1998 to 2004, to account for different exposures to violence as children. This procedure assumes, following Pollak (2002), that this exposure determines current predisposition to tolerate domestic violence. A very large proportion of predicted values fell outside the reported bracket, so we opted instead for interval data estimation using 2004 data only[2]. This measure of income is also less likely to be endogenous. A predicted

[2]Interval data estimation does not account for the simultaneity between violence and income. The 2005 BHPS wave includes personality variables which we could use instead of fixed effects, but the actual questions differ from the questions used in our survey. So we opted for a simple interval data regression.

measure of household income filters out many shocks caused by unobserved factors that might simultaneously influence happiness (Luttmer, 2005). This paper will additionally account for factors which may reduce the simultaneity between income and life satisfaction even further, and which are often omitted from most surveys. These include a distress index which is also likely to capture unexpected shocks to income; personality variables which partly capture the unobserved heterogeneity that explains positive correlations between happiness and income (and between happiness and experienced violence); and a predicted log hourly wage, which according to Pollak (2005), is the appropriate measure to capture outside options in a relationship, and can partially account for social comparisons and reference groups. Regrettably, this survey does not include information on spouses.

3.3.1 Domestic Violence in the UK: a few descriptive results

Appendix F shows a summary of all the variables used in this paper. Table 3.3 shows how the two measures of experienced violence change with the variables used in this paper. It shows that individuals who have been victims of domestic violence earn a lower income, both the victim and the household where s\he lives, and are less happy. It is also quite clear that the measure of violence which captures more recent experiences has a higher proportion of respondents saying they are completely dissatisfied than the coarser measure of violence (even if in actual absolute frequencies, this is a very small number).

Table 3.3: Incidence of domestic violence

	Victim of Domestic Violence			
	Ever		**Recently**	
	Yes	**No**	**Yes**	**No**
Individual Personal Income$^{a)}$				
No income	3.82	5.14	4.92	4.86
£1 up to £9,999 a year	39.49	27.54	39.34	29.36

Continued on next page

	Ever		Recently	
	Yes	**No**	**Yes**	**No**
£10,000 up to £19,999	34.39	31.23	34.43	31.45
£20,000 up to £29,999	15.29	20.55	13.11	20.00
£30,000 up to £39,999	5.10	9.75	6.56	9.02
£40,000 or more a year	1.91	5.80	1.64	5.32
N	157	759	61	865
[a] p-value of Wilcoxon rank-sum equality test	0.001		0.041	
Life Satisfaction[b]				
Completely Dissatisfied	1.15	1.16	5.41	1.03
Very Dissatisfied	11.49	4.07	10.81	4.83
Fairly Dissatisfied	18.97	10.48	24.32	11.40
Neither Satisfied nor Dissatisfied	9.77	9.31	16.22	8.83
Fairly Satisfied	37.36	43.66	29.73	43.43
Very Satisfied	18.97	26.43	10.81	25.98
Completely Satisfied	2.30	4.89	2.70	4.52
N	174	859	74	974
[b] p-value of Wilcoxon rank-sum equality test	0.000		0.000	
Ethnicity[c]				
White British	86.47	91.12	88.57	90.51
Non-White British	13.53	8.88	11.43	9.49
N	170	833	70	948
[c] p-value of Wilcoxon rank-sum equality test	0.062		0.597	
Education[d]				
Other schooling	10.78	8.50	12.12	8.64
Vocational Diploma	34.73	27.09	30.30	28.40
CSE or A Levels	35.93	34.24	39.39	34.02
Graduate	18.56	30.17	18.18	28.94
N	167	812	66	926
[d] p-value of Wilcoxon rank-sum equality test	0.002		0.104	
Marital Status[e]				
Married or co-habiting	59.77	67.17	52.70	66.43
Separated	18.97	6.52	18.92	8.11
Other living alone	21.26	26.31	28.38	25.46

Continued on next page

	Ever		Recently	
	Yes	**No**	**Yes**	**No**
N	174	859	74	974
[e] p-value of Wilcoxon rank-sum equality test	0.343		0.055	

Number of Dependents[f]				
None	61.49	70.43	59.46	69.82
At least one dependent	38.51	29.57	40.54	30.18
N	174	859	74	974
[f] p-value of Wilcoxon rank-sum equality test	0.020		0.063	

Work Status[g]				
Working	52.87	58.91	55.41	57.49
Not working	47.13	41.09	44.59	42.51
N	174	859	74	974
[g] p-value of Wilcoxon rank-sum equality test	0.142		0.726	

Gross Household Income[h]				
no income	no obs	1.12	1.82	0.86
£1 up to £9,999 a year	22.30	10.64	20.00	12.39
£10,000 up to £19,999	31.76	22.69	40.00	23.31
£20,000 up to £29,999	15.54	24.23	14.55	23.19
£30,000 up to £39,999	18.24	18.07	12.73	18.28
$\geq$ £40,000 or more a year	12.16	23.25	10.91	21.96
N	148	714	55	815
[h] p-value of Wilcoxon rank-sum equality test	0.000		0.001	

Psychological Distress[i]				
0	27.59	40.86	21.62	39.73
1	13.22	16.30	10.81	16.02
2	17.82	12.69	12.16	13.55
3	12.64	8.96	8.11	9.65
4	6.90	8.85	12.16	8.32
5	7.47	4.89	10.81	4.93
6	7.47	4.89	17.57	4.62
7	6.90	2.56	6.76	3.18
N	174	859	74	974
[i] p-value of Wilcoxon rank-sum equality test	0.000		0.000	

Continued on next page

	Ever		Recently	
	Yes	**No**	**Yes**	**No**
Predicted Hourly Wage Rate[j)]				
Mean Hourly Rate (/hour)	15.30	15.40	13.35	15.47
N	125	517	35	615
[j)] p-value of Wilcoxon rank-sum equality test	0.991		0.238	
Age[k)]				
Mean age	44.64	44.00	37.89	44.60
N	174	859	74	974
[k)] p-value of Wilcoxon rank-sum equality test	0.641		0.000	
Crime Rate[l)]				
Mean Crime Rate (number of crimes / 1000 people)	27.32	28.35	28.41	28.13
N	157	764	65	869
[l)] p-value of Wilcoxon rank-sum equality test	0.908		0.674	
Personality indicators				
Extraversion[m)]	0.16	0.28	0.26	0.25
Agreeableness[n)]	0.88	0.67	1.03	0.67
Conscientiousness[o)]	0.39	0.29	0.18	0.32
Emotional Stability[p)]	0.35	0.39	0.34	0.38
Openness[q)]	0.14	0.26	0.17	0.24
N	174	859	74	974
[m)] p-value of Wilcoxon rank-sum equality test	0.061		0.868	
[n)] p-value of Wilcoxon rank-sum equality test	0.005		0.002	
[o)] p-value of Wilcoxon rank-sum equality test	0.224		0.183	
[p)] p-value of Wilcoxon rank-sum equality test	0.354		0.825	
[q)] p-value of Wilcoxon rank-sum equality test	0.066		0.360	

While we would expect the average age of victims to be slightly higher, which is true when we look at all experiences of violence (even if not significant), results show that average age of victims of recent events are likely to be younger than the group of non-victims. This is a result that has not been exploited in the literature and may worth further research. The ethnic composition of victims has more non-Whites than the composition of non-victims, which confirms findings stating the incidence of domestic violence is higher amongst Blacks, Asian and minorities, groups for whom data and the analysis of the causes and consequences of domestic violence are even more scarce. This higher incidence of domestic violence amongst non-Whites may reflect characteristics of the households which makes them more vulnerable to domestic violence, and is no longer significant for more recent experiences. For instance, non-White women have on average a larger number of children, and are more likely to have children and the incidence of domestic violence amongst respondents with children is highest than amongst those without children. This is in line with Agarwal (2006), which claims that the number of children deter women more from leaving a violent relationship. Non-whites are also less educated on average and the proportion of graduates who are victims is much lower than the proportion of non-victims.

Marital status is a key variable in this study and reveals somewhat surprising results. Domestic violence can only occur if the respondent lives with the perpetrator, but evidence does suggest that some of the worst cases of domestic violence, by which we mean violence inflicted by a current or former spouse or family member, happen to individuals while a relationship breaks down (and just after). We have grouped individuals according to whether they are living with someone, whether they are separated, or whether they do not have a partner (singles and widowers). While most of violent incidents are perpetrated by partners, there is no reason why individuals without a partner would not be current victims of domestic violence, also because it can be inflicted by other members of the family. The incidence of domestic violence amongst separated respondents is highest than amongst the remaining two groups which may not only reflect the fact that respondents may have terminated an abusive relationship, but also the fact that separation, for whichever reasons, may have generated violence. The proportion of married respondents who are victims of violence is lower than those who are not, and this gap is even larger for the second measure of violence. And surprisingly, while the proportion of singles and widowers in the victims' group of the first measure is lower than in the non-victims' group, this is reversed when we look at the second measure of violence. This is another result about which the literature has little to say. On the other hand, whether or not the respondent is working does not seem to vary systematically with experienced violence. This is even more so for the second measure of experienced violence. Given the low content of this variable, we exclude it from our regression equations.

We follow Pollak (2005)'s suggestion and use the individual predicted wage rate as a measure of the strength of one's threat point. This predicted wage rate was estimated with fixed effects by matching individuals with their BHPS counterparts. Data from 1998 until 2004 was used and the regressors of the wage equation were gender, age and age squared, the calendar year, educational attainment, employment status, marital status, number of dependents, ethnicity, religion and fine regional data (and so it excludes direct information on experienced violence which is not captured by income or fixed effects). This measure can also be thought of as a measure of relative income and social comparisons. Table 3.3 shows that victims of domestic violence do not seem to have different predicted wages from non-victims, when we look at the first measure of violence. However, the difference between victims and non-victims increases with the second measure, but is still not statistically significant.

There is evidence that certain personality traits such as being sympathetic or not being quarrelsome are highly correlated with the presence of domestic violence. Lundberg (2010) shows that individuals with certain personality traits, such as agreeableness, are less likely to divorce. Pollak (2002) shows that under plausible assumptions, there is also a persistent intergenerational impact of domestic violence, which is partly determined by intergenerational transmission of personality and upbringing. Based on Gosling et al. (2003), the dataset includes the following ten questions on individual personality traits:

1. I see myself as extraverted, enthusiastic (7 point scale: 1=agree strongly 7=disagree strongly)
2. I see myself as reserved, quiet (7 point scale: 1=agree strongly 7=disagree strongly)
3. I see myself as sympathetic, warm (7 point scale: 1=agree strongly 7=disagree strongly)
4. I see myself as critical, quarrelsome (7 point scale: 1=agree strongly 7=disagree strongly)
5. I see myself as dependable, self-disciplined (7 point scale: 1=agree strongly 7=disagree strongly)
6. I see myself as disorganised, careless (7 point scale: 1=agree strongly 7=disagree strongly)
7. I see myself as calm, emotionally stable (7 point scale: 1=agree strongly 7=disagree strongly)
8. I see myself as anxious, easily upset (7 point scale: 1=agree strongly 7=disagree strongly)
9. I see myself as open to new experience, complex (7 point scale: 1=agree strongly 7=disagree strongly)
10. I see myself as conventional, uncreative (7 point scale: 1=agree strongly 7=disagree strongly)

These 10 traits give rise to 5 personality dimensions. Extraversion is the combination of the first two polarised traits, i.e. extraverted and reserved. The negative trait is given a negative sign and the two are averaged to yield extraversion. The remaining 4 dimensions result from a similar averaging of two opposite traits, and yield agreeableness, conscientiousness, emotional stability and openness. Therefore, each personality variable takes values from -6 to 6. The table shows the average value for each personality dimension of both groups of respondents, for both measures. Victims tend to be more agreeable, confirming the result already found in Lundberg (2010). Victims also tend to be more psychologically distressed, and this is also true and very significant when we look at the first measure of violence. The index of psychological distress partially captures some of the impact of violence on life satisfaction we would want to estimate. However, given that omitting this variable may amplify the bias in the coefficient of income, we run regressions with and without this variable and present both sets of results.

As suggested in Morrison & Biehl (1999), higher violent crime rates lower inhibitions against violent conduct, both via a demonstration effect (emulation of violent behaviour) and via erosion of social norms that regulate interpersonal relations. This data set also includes each individual 3-digit postcode, which we use to match each individual to the crime rate in their neighbourhood (as in Anand & Santos, 2007). Local crime data were collected online from `http://www.crimestatistics.org.uk/tool/`. This variable measures the number of all reported crime offences per 1000 individuals in the first quarter of 2004. This rate includes all types of assault and not just the bodily harm offences and was chosen to prevent arbitrary assumptions about which subcategories of assault do not contribute to erode norms about violence. There is no significant difference in the average crime rate between victims and non-victims for both measures of violence.

3.4 Estimating a Utility function

This paper estimates the costs of domestic violence as the compensating variation needed to compensate an individual for having experienced violence. The basic life satisfaction equation and parameter of interest were discussed in Section 3.2, and summarised in Eqs. 3.1 and 3.2. We next discuss the functional forms chosen for household income $f(y_i)$ and consequently, for our parameter of interest, CV.

3.4.1 Specifying the relationship between income and utility

While the linear or the logarithmic functions have been used more often, these income specifications impose strong assumptions on how compensating variation varies with income. Our preferred specification models income as a quadratic function, as follows:

$$E\left(U_i \middle| DV_i, y_i, X_i\right) = \alpha_0 + \alpha_1 DV_i + \alpha_2 y_i + \alpha_3 y_i^2 + \alpha' X_i + \varepsilon_i \tag{3.3}$$

CV calculated as Eq. 3.4 shows:

$$CV = -\left(y_0 + \frac{\alpha_2}{2\alpha_3}\right)\left[1 \pm \sqrt{1 - \frac{\alpha_1}{\alpha_3\left(y_0 + \frac{\alpha_2}{2\alpha_3}\right)^2}}\right] \tag{3.4}$$

Conditional on the parameters α_1, α_2 and α_3, one root will represent an increasing relation between income and CV, and the other root will represent the opposite relation. We will nevertheless always choose the minimum root, because CV is the *minimum* amount of income needed to compensate victims[3]. The relation with income will remain dependent on the parameters and initial income level.

We have made attempts to use a Box-Cox function as an alternative specification, and to test for the strength of the linear and the logarithmic specifications of income, often used in the literature. The estimation results using the male sample only are very imprecise, which decreases the power of the tests. For the female sample, we seem to find some support for the linear specification using the point estimate of income measure. When we move to imputed income, this evidence becomes much stronger. Broadly, when we move to the second measure of income, the predicted

[3]Unless the lowest root is negative. This may be the case when $\alpha_3 > 0$.

income values, standard errors and statistical significance improve and shows us that the linear specification continues to receive stronger support for the female sample, and the λ point estimate is now much closer to 1. All estimates of λ are not larger than 1, which suggests an independent or increasing relation between CV and income. The following table summarises the results obtained for the Box Cox specification of income[4].

Table 3.4: Testing for alternative specifications of income in the life satisfaction equation: results from Box Cox specification, parameter λ

	Domestic Violence ever			**Domestic Violence recently**		
	All	**Women**	**Men**	**All**	**Women**	**Men**
Using point estimate of household income						
λ coefficient	1.457	2.924	0.108	1.318	2.571	0.118
λ standard error	(0.766)	(1.705)	(0.589)	(0.747)	(1.599)	(0.621)
p-value $H_0 : \lambda = 0$	0.008	0.001	0.852	0.018	0.004	0.847
p-value $H_0 : \lambda = 1$	0.497	0.066	0.210	0.642	0.140	0.242
Using predicted household income						
λ coefficient	0.429	0.986	-0.143	0.389	0.923	-0.162
λ standard error	(0.297)	(0.430)	(0.360)	(0.303)	(0.445)	(0.371)
p-value $H_0 : \lambda = 0$	0.147	0.026	0.690	0.199	0.047	0.662
p-value $H_0 : \lambda = 1$	0.085	0.975	0.009	0.071	0.864	0.010

These results seem to suggest a linear specification is appropriate for the female sample. When looking at the whole sample, results are also consistent with a logarithmic specification of income. Given that the first measure of income only has six distinct points, we think the results of the second measure may be more informative, despite the overestimation of victims' income. But given the weakness of the evidence, and how it varies with the income measure and regression used, we will present our estimation results using the quadratic function of income, a linear

[4]When the distress index was removed, and given the negative impact it has on happiness and the negative correlation it has with household income (around -0.08), estimates decreased. Results also became less precise and the power of the tests decreased slightly. Qualitatively results did not change though and will not be presented here.

in parameters model which requires lower demands on the income data.

3.4.2 Estimation results of life satisfaction equation

Table 3.5 shows the estimation results of life satisfaction equations defined according to Eq. 3.3. We present results for both measures of income (income point estimates in the first 6 columns, and predicted income for the last 6 columns), and for both measures of experienced violence (recent experiences in columns 4-6 and 10-12, and experiences over entire life span in columns 1-3 and 7-9). And we estimate our life satisfaction equation for the whole sample (columns 1, 4, 7, 10), women (columns 2, 5, 8, 11) and men (columns 3, 6, 9, 12).

The socio-demographic indicators used are explained in Appendix F. We include a gender dummy, a quadratic function of age, marital status, ethnicity, presence of dependents and education. To account for the endogeneity of domestic violence and the self-selection of victims into abusive relations, we use personality indicators, and a measure of outside options, the log hourly wage predicted from BHPS. We also use local crime rates which account for norms related to violence and quality of public services. To account for the endogeneity of household income, we also use the distress index, which should reflect the shocks to utility that lead individuals to revise their income generating decisions. Potential wage can partly also account for the importance social comparisons and reference groups have for individuals. However, because distress partly captures the impact of violence this study aims to estimate, it is likely that its inclusion underestimates the effects of violence on life satisfaction. We also exclude this variable to assess how results change in Table 3.6.

Table 3.5: Happiness equations: main estimation results

	Income: Point estimate						Income: Predicted income					
	Domestic Violence ever			Domestic Violence recently			Domestic Violence ever			Domestic Violence recently		
	All	**Women**	**Men**	**All**	**Women**	**Men**	**All**	**Women**	**Men**	**All**	**Women**	**Men**
Domestic Violence	-0.254**	-0.303**	-0.179	-0.581***	-0.624**	-0.542*	-0.228**	-0.257*	-0.169	-0.526***	-0.546**	-0.544*
	(0.115)	(0.142)	(0.202)	(0.193)	(0.246)	(0.325)	(0.114)	(0.143)	(0.195)	(0.191)	(0.245)	(0.322)
Household income/10000	0.002	-0.228	0.360*	0.018	-0.177	0.328*	0.219***	0.117	0.342***	0.216***	0.120	0.336***
	(0.119)	(0.158)	(0.194)	(0.118)	(0.155)	(0.194)	(0.065)	(0.095)	(0.094)	(0.064)	(0.094)	(0.094)
Household income/10000^2	0.021	0.055**	-0.027	0.018	0.046**	-0.023	-0.011**	-0.000	-0.022***	-0.011**	-0.001	-0.021***
	(0.016)	(0.022)	(0.026)	(0.016)	(0.021)	(0.026)	(0.006)	(0.009)	(0.008)	(0.006)	(0.009)	(0.008)
Female	0.188**			0.157*			0.220**			0.194**		
	(0.089)			(0.088)			(0.089)			(0.089)		
age	-0.164***	-0.146**	-0.177**	-0.164***	-0.152**	-0.170**	-0.171***	-0.158**	-0.176**	-0.172***	-0.164***	-0.169**
	(0.044)	(0.061)	(0.070)	(0.043)	(0.059)	(0.070)	(0.044)	(0.061)	(0.069)	(0.043)	(0.060)	(0.069)
age^2	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Separated	-0.078	-0.156	-0.046	-0.082	-0.157	-0.067	-0.020	-0.097	-0.028	-0.021	-0.099	-0.042
	(0.152)	(0.203)	(0.241)	(0.148)	(0.198)	(0.236)	(0.152)	(0.206)	(0.238)	(0.148)	(0.201)	(0.234)
No partner	-0.363***	-0.380**	-0.349*	-0.382***	-0.391**	-0.381*	-0.336***	-0.368**	-0.318	-0.355***	-0.383**	-0.348*
	(0.130)	(0.173)	(0.210)	(0.129)	(0.172)	(0.210)	(0.130)	(0.176)	(0.208)	(0.129)	(0.174)	(0.207)
Non-White British	-0.108	-0.018	-0.181	-0.146	-0.050	-0.226	-0.121	-0.050	-0.213	-0.144	-0.054	-0.258
	(0.164)	(0.220)	(0.257)	(0.163)	(0.220)	(0.255)	(0.162)	(0.220)	(0.253)	(0.161)	(0.218)	(0.251)
At least 1 child	0.146	0.225	0.062	0.182*	0.278*	0.074	0.154	0.238	0.058	0.190*	0.287*	0.071
	(0.105)	(0.152)	(0.154)	(0.104)	(0.151)	(0.153)	(0.104)	(0.154)	(0.152)	(0.104)	(0.152)	(0.151)
Vocational diploma	0.073	0.314	-0.261	0.090	0.353	-0.278	0.056	0.314	-0.273	0.071	0.348	-0.297
	(0.161)	(0.234)	(0.231)	(0.160)	(0.231)	(0.230)	(0.161)	(0.236)	(0.228)	(0.160)	(0.232)	(0.227)
CSE A level	0.125	0.180	-0.039	0.137	0.191	-0.030	0.113	0.192	-0.051	0.120	0.200	-0.050

Continued on next page

	Income: Point estimate						Income: Predicted income					
	Domestic Violence ever			Domestic Violence recently			Domestic Violence ever			Domestic Violence recently		
	All	**Women**	**Men**	**All**	**Women**	**Men**	**All**	**Women**	**Men**	**All**	**Women**	**Men**
	(0.161)	(0.226)	(0.241)	(0.159)	(0.222)	(0.240)	(0.161)	(0.227)	(0.238)	(0.159)	(0.223)	(0.236)
Graduate	0.118	0.181	-0.025	0.147	0.237	-0.047	0.123	0.204	-0.031	0.148	0.255	-0.055
	(0.167)	(0.239)	(0.242)	(0.165)	(0.235)	(0.242)	(0.166)	(0.239)	(0.240)	(0.165)	(0.235)	(0.239)
Extraversion	0.191***	0.117	0.279***	0.190***	0.120	0.285***	0.182***	0.100	0.277***	0.184***	0.102	0.287***
	(0.059)	(0.081)	(0.088)	(0.058)	(0.079)	(0.087)	(0.059)	(0.082)	(0.086)	(0.058)	(0.080)	(0.085)
Agreeableness	-0.022	-0.039	-0.002	-0.021	-0.029	-0.011	-0.031	-0.057	-0.001	-0.028	-0.041	-0.009
	(0.049)	(0.070)	(0.071)	(0.049)	(0.069)	(0.070)	(0.049)	(0.070)	(0.070)	(0.049)	(0.070)	(0.069)
Conscientiousness	0.154***	0.162**	0.120	0.152***	0.153**	0.123	0.157***	0.189***	0.117	0.153***	0.176**	0.118
	(0.052)	(0.072)	(0.076)	(0.051)	(0.070)	(0.075)	(0.051)	(0.073)	(0.075)	(0.051)	(0.071)	(0.074)
Emotional stability	-0.023	0.011	-0.019	-0.027	0.000	-0.012	-0.007	0.029	-0.018	-0.009	0.026	-0.010
	(0.054)	(0.075)	(0.079)	(0.053)	(0.075)	(0.079)	(0.054)	(0.075)	(0.079)	(0.053)	(0.075)	(0.079)
Openness	0.061	0.166**	-0.075	0.072	0.173**	-0.063	0.058	0.159**	-0.079	0.067	0.164**	-0.067
	(0.049)	(0.068)	(0.074)	(0.049)	(0.067)	(0.074)	(0.049)	(0.069)	(0.074)	(0.049)	(0.068)	(0.074)
Predicted log hourly wage	0.244*	0.109	0.397*	0.206*	0.057	0.374*	0.247*	0.131	0.389*	0.216*	0.086	0.363*
	(0.128)	(0.163)	(0.221)	(0.125)	(0.157)	(0.221)	(0.128)	(0.166)	(0.218)	(0.125)	(0.160)	(0.218)
Distressed	-0.285***	-0.309***	-0.270***	-0.282***	-0.306***	-0.265***	-0.283***	-0.305***	-0.272***	-0.280***	-0.303***	-0.267***
	(0.022)	(0.030)	(0.034)	(0.022)	(0.030)	(0.034)	(0.022)	(0.030)	(0.033)	(0.022)	(0.030)	(0.033)
Local crime rates	0.003	0.001	0.003	0.003	0.000	0.003	0.004	0.003	0.003	0.003	0.002	0.003
	(0.003)	(0.006)	(0.003)	(0.003)	(0.006)	(0.003)	(0.003)	(0.006)	(0.003)	(0.003)	(0.006)	(0.003)
Constant	7.993***	8.108***	7.861***	7.977***	8.178***	7.753***	7.842***	7.897***	7.842***	7.871***	7.999***	7.707***
	(0.975)	(1.322)	(1.544)	(0.959)	(1.295)	(1.540)	(0.973)	(1.339)	(1.513)	(0.957)	(1.308)	(1.508)
R^2	0.368	0.371	0.369	0.379	0.387	0.374	0.360	0.345	0.378	0.372	0.365	0.383
N	577	310	267	585	318	267	586	316	270	594	324	270

Significance levels : * 10% ** 5% *** 1% Standard errors in parentheses

Omitted categories: being a man, married, other schooling, lower relative income, White British, no dependents. Standard errors of income multiplied by 10000.

Using either measure, domestic violence has a significantly pervasive impact on life satisfaction for women. The effect is not significant for men with the first measure, but becomes as large and significant when violence only measures recent events. So despite the fact that the nature of the violence suffered by men may be different and often leads to different behavioural responses which could be argued to lead to adaptation and undermine this estimation exercise, evidence seems to suggest this is not the case when the violence measure relates to recent events. Results do not vary greatly with the measure of income used. When we exclude the distress index, magnitudes of the impact of violence do increase, but significance patterns remain the same.

Gross Household income increases life satisfaction at a decreasing rate for men, but the pattern that seems to be emerging for women is much less precise. When using the income point estimate, and for both measures of violence, we obtain a convex relation with increasing returns, but this pattern disappears for the imputed income measure. When we include distress in Table 3.5, there does not seem any relation at all between income and life satisfaction, but when we exclude the distress index in Table 3.6, the estimated relationship is linear. This weaker impact of income on women's assessments of current life than men's assessment is already documented in the literature, but it does raise concerns about the validity of this methodology. We will present the estimates of the costs of domestic violence only for the regression equations where this impact was significant.

Table 3.6: Happiness equations: excluding distress index

	Income: Point estimate						Income: Predicted income					
	Domestic Violence ever			Domestic Violence recently			Domestic Violence ever			Domestic Violence recently		
	All	**Women**	**Men**	**All**	**Women**	**Men**	**All**	**Women**	**Men**	**All**	**Women**	**Men**
Domestic Violence	-0.377***	-0.381**	-0.357	-1.018***	-1.131***	-0.903**	-0.337***	-0.320*	-0.329	-0.949***	-1.019***	-0.924**
	(0.131)	(0.166)	(0.225)	(0.216)	(0.280)	(0.360)	(0.129)	(0.166)	(0.218)	(0.214)	(0.277)	(0.357)
Household income/10000	0.113	-0.110	0.476**	0.138	-0.033	0.420*	0.256***	0.188*	0.346***	0.259***	0.204*	0.335***
	(0.136)	(0.184)	(0.217)	(0.134)	(0.180)	(0.217)	(0.074)	(0.110)	(0.106)	(0.073)	(0.108)	(0.105)
Household income/10000^2	0.008	0.044*	-0.043	0.004	0.032	-0.036	-0.014**	-0.004	-0.022**	-0.014**	-0.006	-0.021**
	(0.018)	(0.025)	(0.029)	(0.018)	(0.025)	(0.029)	(0.006)	(0.011)	(0.009)	(0.006)	(0.010)	(0.009)
Female	0.230**			0.187*			0.258**			0.222**		
	(0.102)			(0.100)			(0.102)			(0.100)		
age	-0.179***	-0.159**	-0.209***	-0.182***	-0.174**	-0.195**	-0.186***	-0.175**	-0.204***	-0.190***	-0.189***	-0.190**
	(0.050)	(0.071)	(0.078)	(0.049)	(0.069)	(0.078)	(0.050)	(0.071)	(0.078)	(0.049)	(0.069)	(0.077)
age^2	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***	0.002***
	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)
Separated	-0.088	-0.072	-0.166	-0.069	-0.014	-0.212	-0.044	-0.014	-0.176	-0.014	0.049	-0.203
	(0.174)	(0.237)	(0.270)	(0.169)	(0.230)	(0.263)	(0.173)	(0.239)	(0.267)	(0.168)	(0.232)	(0.261)
No partner	-0.422***	-0.287	-0.557**	-0.453***	-0.308	-0.607***	-0.390***	-0.271	-0.529**	-0.420***	-0.294	-0.576**
	(0.148)	(0.202)	(0.233)	(0.147)	(0.200)	(0.232)	(0.147)	(0.203)	(0.232)	(0.146)	(0.201)	(0.230)
Non-White British	-0.100	0.005	-0.214	-0.163	-0.044	-0.296	-0.100	-0.018	-0.244	-0.136	-0.020	-0.324
	(0.187)	(0.258)	(0.288)	(0.185)	(0.256)	(0.284)	(0.184)	(0.255)	(0.285)	(0.183)	(0.252)	(0.281)
At least 1 child	0.176	0.252	0.060	0.228*	0.324*	0.076	0.183	0.264	0.051	0.236**	0.337*	0.073
	(0.119)	(0.178)	(0.172)	(0.118)	(0.176)	(0.171)	(0.119)	(0.178)	(0.171)	(0.118)	(0.176)	(0.169)
Vocational diploma	0.073	0.158	-0.141	0.089	0.218	-0.169	0.056	0.138	-0.125	0.069	0.193	-0.170
	(0.184)	(0.273)	(0.258)	(0.182)	(0.268)	(0.257)	(0.184)	(0.273)	(0.256)	(0.181)	(0.268)	(0.254)
CSE A level	0.135	0.147	-0.017	0.137	0.148	0.002	0.111	0.120	-0.004	0.105	0.114	-0.001

Continued on next page

	Income: Point estimate						Income: Predicted income					
	Domestic Violence ever			Domestic Violence recently			Domestic Violence ever			Domestic Violence recently		
	All	**Women**	**Men**	**All**	**Women**	**Men**	**All**	**Women**	**Men**	**All**	**Women**	**Men**
	(0.184)	(0.265)	(0.270)	(0.181)	(0.258)	(0.267)	(0.183)	(0.264)	(0.267)	(0.180)	(0.257)	(0.265)
Graduate	0.101	0.100	0.001	0.125	0.165	-0.035	0.097	0.096	0.010	0.115	0.152	-0.033
	(0.190)	(0.279)	(0.271)	(0.187)	(0.273)	(0.270)	(0.189)	(0.277)	(0.270)	(0.187)	(0.272)	(0.268)
Extraversion	0.324***	0.273***	0.372***	0.317***	0.258***	0.383***	0.321***	0.267***	0.369***	0.316***	0.251***	0.384***
	(0.066)	(0.093)	(0.098)	(0.065)	(0.091)	(0.097)	(0.065)	(0.093)	(0.096)	(0.064)	(0.091)	(0.095)
Agreeableness	-0.095*	-0.106	-0.085	-0.086	-0.073	-0.100	-0.105*	-0.121	-0.089	-0.092*	-0.083	-0.101
	(0.056)	(0.081)	(0.079)	(0.055)	(0.080)	(0.077)	(0.055)	(0.081)	(0.078)	(0.055)	(0.080)	(0.077)
Conscientiousness	0.114*	0.134	0.071	0.102*	0.102	0.077	0.120**	0.166*	0.076	0.106*	0.129	0.077
	(0.059)	(0.084)	(0.085)	(0.058)	(0.082)	(0.084)	(0.058)	(0.084)	(0.084)	(0.057)	(0.082)	(0.083)
Emotional stability	-0.025	-0.012	-0.001	-0.032	-0.032	0.011	-0.012	-0.000	-0.003	-0.015	-0.006	0.010
	(0.061)	(0.088)	(0.089)	(0.061)	(0.088)	(0.088)	(0.061)	(0.087)	(0.089)	(0.061)	(0.087)	(0.088)
Openness	0.061	0.127	-0.034	0.080	0.142*	-0.014	0.061	0.127	-0.036	0.078	0.139*	-0.016
	(0.056)	(0.079)	(0.083)	(0.056)	(0.078)	(0.083)	(0.056)	(0.080)	(0.083)	(0.056)	(0.078)	(0.082)
Predicted log hourly wage	0.277*	0.150	0.479*	0.238*	0.109	0.439*	0.277*	0.180	0.463*	0.248*	0.145	0.419*
	(0.146)	(0.191)	(0.248)	(0.141)	(0.183)	(0.247)	(0.146)	(0.192)	(0.245)	(0.141)	(0.184)	(0.244)
Local crime rates	0.004	0.006	0.003	0.004	0.005	0.004	0.005	0.008	0.003	0.005	0.007	0.003
	(0.003)	(0.007)	(0.004)	(0.003)	(0.007)	(0.004)	(0.003)	(0.007)	(0.004)	(0.003)	(0.007)	(0.004)
Constant	7.496***	7.441***	7.824***	7.529***	7.672***	7.643***	7.457***	7.420***	7.869***	7.542***	7.664***	7.640***
	(1.111)	(1.543)	(1.730)	(1.089)	(1.504)	(1.719)	(1.106)	(1.552)	(1.702)	(1.084)	(1.511)	(1.690)
R^2	0.179	0.141	0.208	0.199	0.171	0.220	0.172	0.120	0.212	0.192	0.152	0.226
N	577	310	267	585	318	267	586	316	270	594	324	270

Significance levels : * 10% ** 5% *** 1% Standard errors in parentheses

Omitted categories: being a man, married, other schooling, lower relative income, White British, no dependents. Standard errors of income multiplied by 10000.

As observed in so many previous studies, women are happier than men, the age-happiness profile is U-shaped, and having no partner is the worst marital status on average (exception is for the female equations estimated without the distress index). Having dependents increases life satisfaction for women on average, and this is never significant for men. Education does not seem to have a significant impact on happiness, also in line with previous studies. Personality does have a significant impact on life satisfaction, specially extroversion and conscientiousness. More extrovert and more conscientious respondents report higher life satisfaction. Potential wage is very significant and contributes to higher life satisfaction for men, but again these results are not significant for women. being distressed does have a major impact on life satisfaction for all 12 equations in Table 3.5, similar in magnitude to the detrimental effects of violence itself. However, excluding this variable, even if it increased the effect of violence, did not substantially alter the remaining coefficients, not even the income coefficients to a large extent.

3.5 Estimating the costs of domestic violence at the individual and at the national level

Table 3.7 shows our estimates of the compensating variation of domestic violence according to Eq. 3.4. These will not be calculated when the impact of income is not significant. When the quadratic term is not significant, CV is estimated according to a linear model.

Costs of DV vary from as little as £1000 to over £50000. When the life satisfaction equation exhibits an increasing but concave profile of income, estimates of CV tend to be lower than when the quadratic term is not significant, which does suggest there is still some work to be done for extreme income values. While the linear model may be overestimating the impact of income changes at extreme values, the quadratic function may be overfitting and underestimating it. These results are not far off from the results obtained in Atkinson et al. (2005).

Table 3.7: Happiness equations: excluding distress index

	Income: Point estimate						Income: Predicted income					
	Domestic Violence ever			Domestic Violence recently			Domestic Violence ever			Domestic Violence recently		
	All	**Women**	**Men**	**All**	**Women**	**Men**	**All**	**Women**	**Men**	**All**	**Women**	**Men**
Main regression												
Individual costs (£)	0	3015.59	4961.74	0	6127.93	16532.76	8621.97	0	1967.62	14148.25	0	4075.45
National costs (*million£*)	0	22688.7	13170.8	0	46105.3	43885.8	36660.9	0	2600.4	60158.9	0	5386.0
% of GDP	0	1.853	1.075	0	3.765	3.583	2.993	0	0.212	4.912	0	0.440
Excluding distress index												
Individual costs (£)	0	4263.03	7504.75	0	0	21508.60	9244.71	17010.10	3107.58	4544.27	50013.52	954.32
National costs (*million£*)	0	32074.2	19921.2	0	0	57094.1	39308.8	49847.3	4106.9	19322.4	146562.4	1261.2
% of GDP	0	2.619	1.627	0	0	4.662	3.210	4.070	0.335	1.578	11.967	0.103

To compute a national estimate of the costs of domestic violence, the first step was to estimate the number of victims in the UK in 2005, the year this questionnaire was delivered. We used the proportion of victims in our sample as a measure of incidence of domestic violence in the whole of the UK (Scotland included). Then we used the estimates of the UK population available in Dye & Sosimi (2006) (over 60 million in 2005, p. 40) to find the number of victims of domestic violence in the UK in 2005. National costs of domestic violence are calculated as the product of number of victims and the individual costs of DV (Table 3.7, row 1). Results are in the second row of the table and show national costs as high as £billion 146, but for the whole sample, never exceeding £billion 39. With a GDP of approximately £million 1,224,715 for the whole UK in 2005 (Dye & Sosimi, 2006, , p. 23), this estimate represents over 3% of the national GDP, in line with the percentage suggested in Walby (2004). Surprisingly, and because the percentage of individuals who feel vulnerable to domestic violence is lower than the percentage of individuals who have ever experienced domestic violence, the estimates of the national costs of DV are similar for the two measures of DV.

3.6 Conclusion

This paper provides an estimate of the costs of domestic violence at the individual and at the aggregate level. It uses a life satisfaction equation where compensating variation is a function of the coefficients of income and domestic violence. It draws on a survey that includes data on experienced violence, household gross income and a self-reported life satisfaction variable. The analysis is conditional on socio-demographic characteristics, potential wage, a distress index, personality and local crime rates. We use personality indicators, potential wage and local crime rates to account for the endogeneity of domestic violence and the self-selection of victims into abusive relations. To account for the endogeneity of household income, we also use the distress index, which should reflect the shocks to utility that lead individuals to revise their income generating decisions.

This paper shows that a satisfaction approach produces estimates which are in line with estimates produced using stated preference methods, as in Atkinson et al. (2005). However, when we use a measure of domestic violence which aims to represent current exposure to domestic violence, we obtain higher individual costs than other studies. In the end, our results confirm that domestic violence is a major inhibitor of individual and social welfare, and compensating victims of domestic violence would cost the UK government up to over 3% of its annual GDP. It is worth emphasising the sensitivity of our estimates to the gender of the respondent, and the sensitivity of the self-reported satisfaction variable to numerous influences. This approach is limited by the possibility that either violence or income not being a substantial part of each respondent's satisfaction. However, it overcomes fundamental limitations of other valuation methods, such as the need to have relevant markets in equilibrium and the incentive to reply strategically. In particular, given that most of the costs of domestic violence are held in private, and are likely to be emotional and human costs for which there are no relevant markets, this approach is, in our view, worth exploring further.

At the same time, there are still reasons to believe that the marginal disutility of violence is underestimated. Self-reported satisfaction will fail to capture the cost of public goods which are unperceived or not valued by the individual or the intergenerational effects of domestic violence, so this measure only captures the costs of domestic violence perceived and understood by the victims. This paper however invites an integrated cost-benefit analysis of domestic violence which takes satisfaction approaches to valuing non-market goods seriously, and shows how urgent this may be for a clearer assessment of the true impact of domestic violence and for a stronger effective support of families where domestic violence occurs.

Appendix A

Comparing the PSID with the CPS: wage distribution and average hours of work over the lifecycle

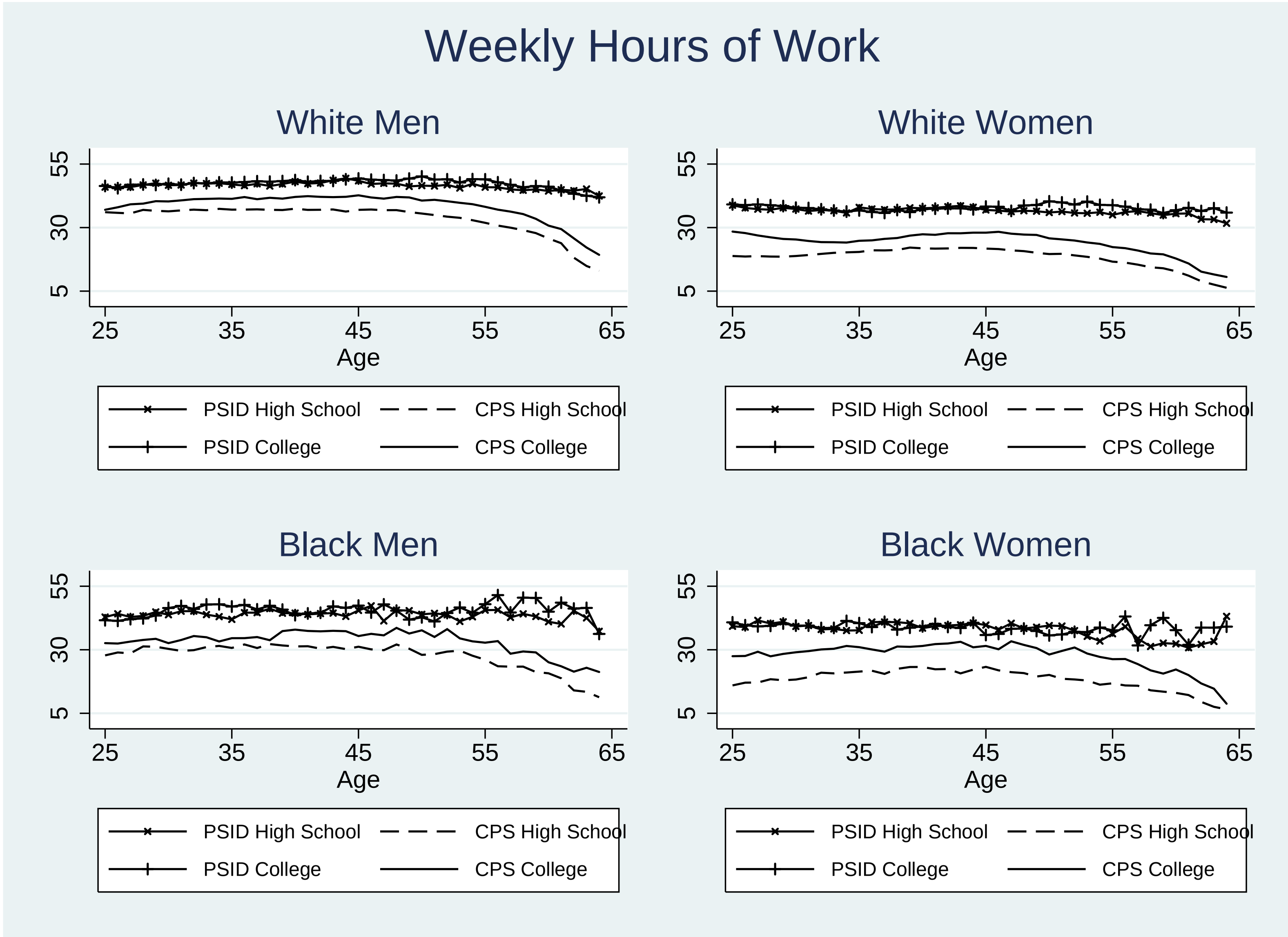


Figure 1: Weekly Hours of Work: PSID and CPS

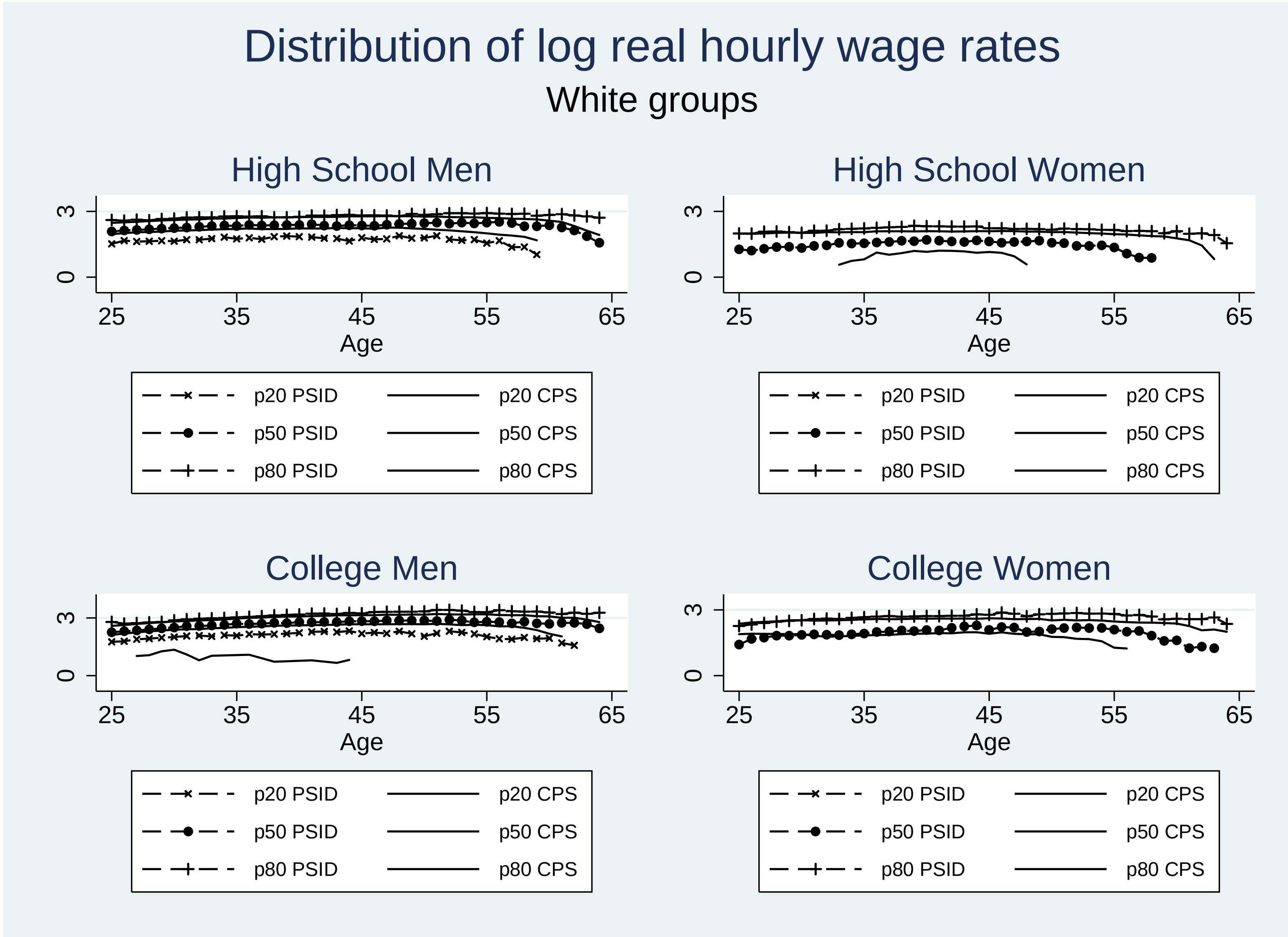


Figure 2: Earnings percentiles by gender and skill group: Whites

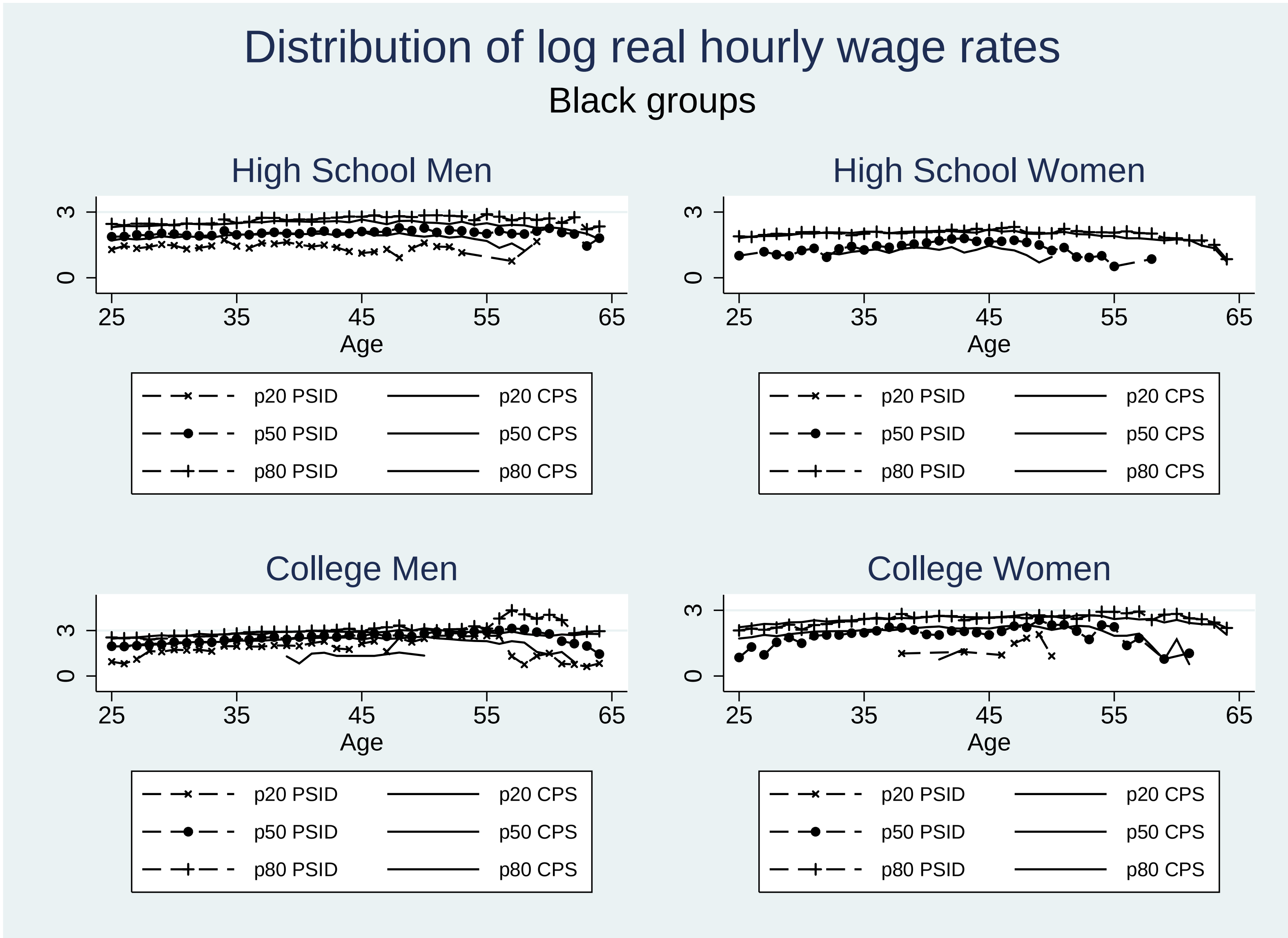


Figure 3: Earnings percentiles by gender and skill group: Blacks

Appendix B

Black male and female average fixed effects over the lifecourse

Section 1.3.2 showed how average fixed effects between White workers and the total sample were not very different, and so selection into employment, albeit significant when comparing workers and non-workers average fixed effects, does not have a strong impact. These profiles were computed using a different set of individuals for different ages, either because individuals dropped the panel, or because the age range we are trying to estimate is longer than the sample period and only individuals from different cohorts can be used to build the whole profile. This appendix shows the figures for the Black groups which demonstrate the relative importance of selection and attrition we analysed in the paper for the White groups. Qualitatively, the results are the same, even if for College groups, attritors do not seem to be significantly different from observed individuals.

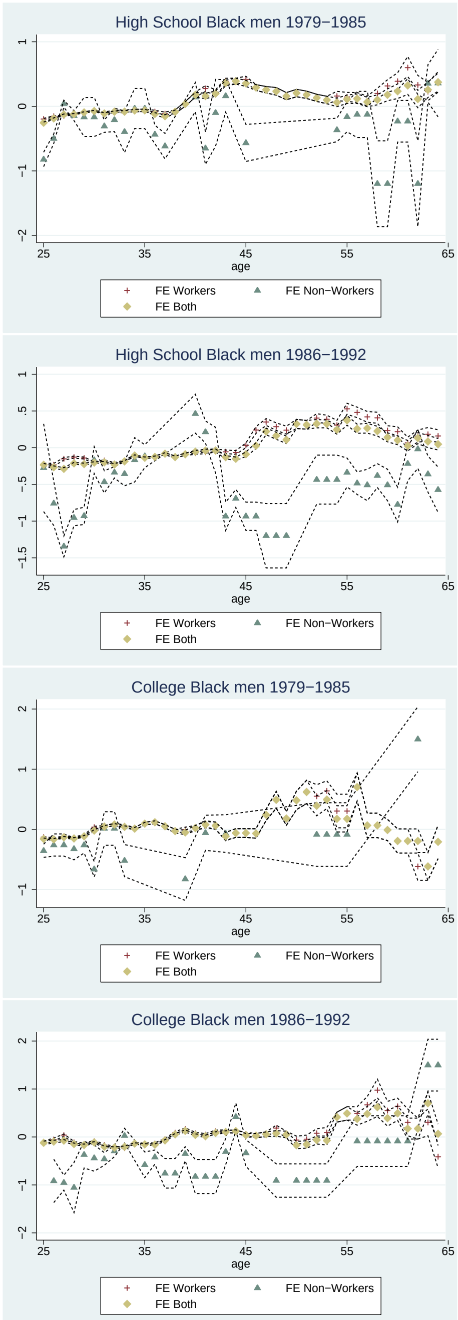


Figure 4: How different are workers from non-workers: Fixed effects of Black men

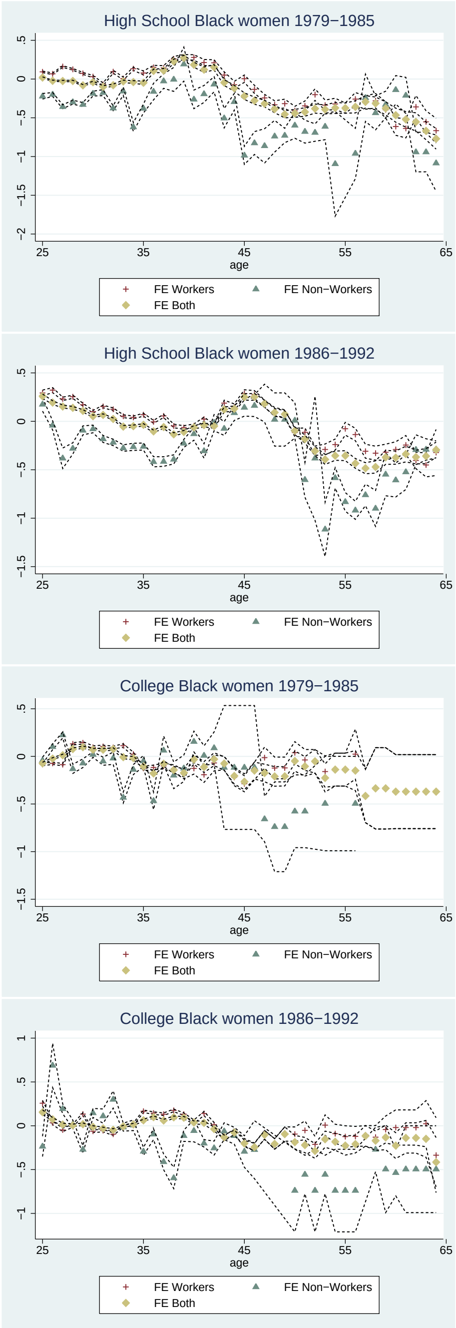


Figure 5: How different are workers from non-workers: Fixed effects of Black women

Table 8: Testing for selective employment and attrition: Blacks

	$H_0 : \overline{FE}_w - \overline{FE}_{Nw} = 0$		$H_0 : \overline{FE}_{\text{obs}} - \overline{FE}_{\text{atr}} = 0$		$H_0 : \overline{FE}_w - \overline{FE}_{\text{exc.w}} = 0$	
	χ^2	***p*-value**	χ^2	***p*-value**	χ^2	***p*-value**
High School Black men						
1979-1985	233.1	0	156.8	0	289.5	0
1986-1992	308.1	0	237.7	0	359.2	0
College Black men						
1979-1985	123.4	0	54.9	0.06	177.8	0
1986-1992	230.6	0	187.5	0	277.1	0
High School Black women						
1979-1985	401.2	0	102.8	0	449.5	0
1986-1992	276.4	0	213.1	0	295.7	0
College Black women						
1979-1985	242.4	0	24.5	0.97	272.6	0
1986-1992	209.2	0	91.2	0	212.1	0

Results based on the χ^2 test $(R\varphi)\,(R\text{CovR}')^{-1}\,(R\varphi)' \sim \chi^2_q$, where φ is the vector of estimated fixed effects
R is the projection matrix that transforms the fixed effects into a difference in means between two groups
Cov is the covariance matrix of the fixed effects
and q is the number of restrictions, one for each age level.

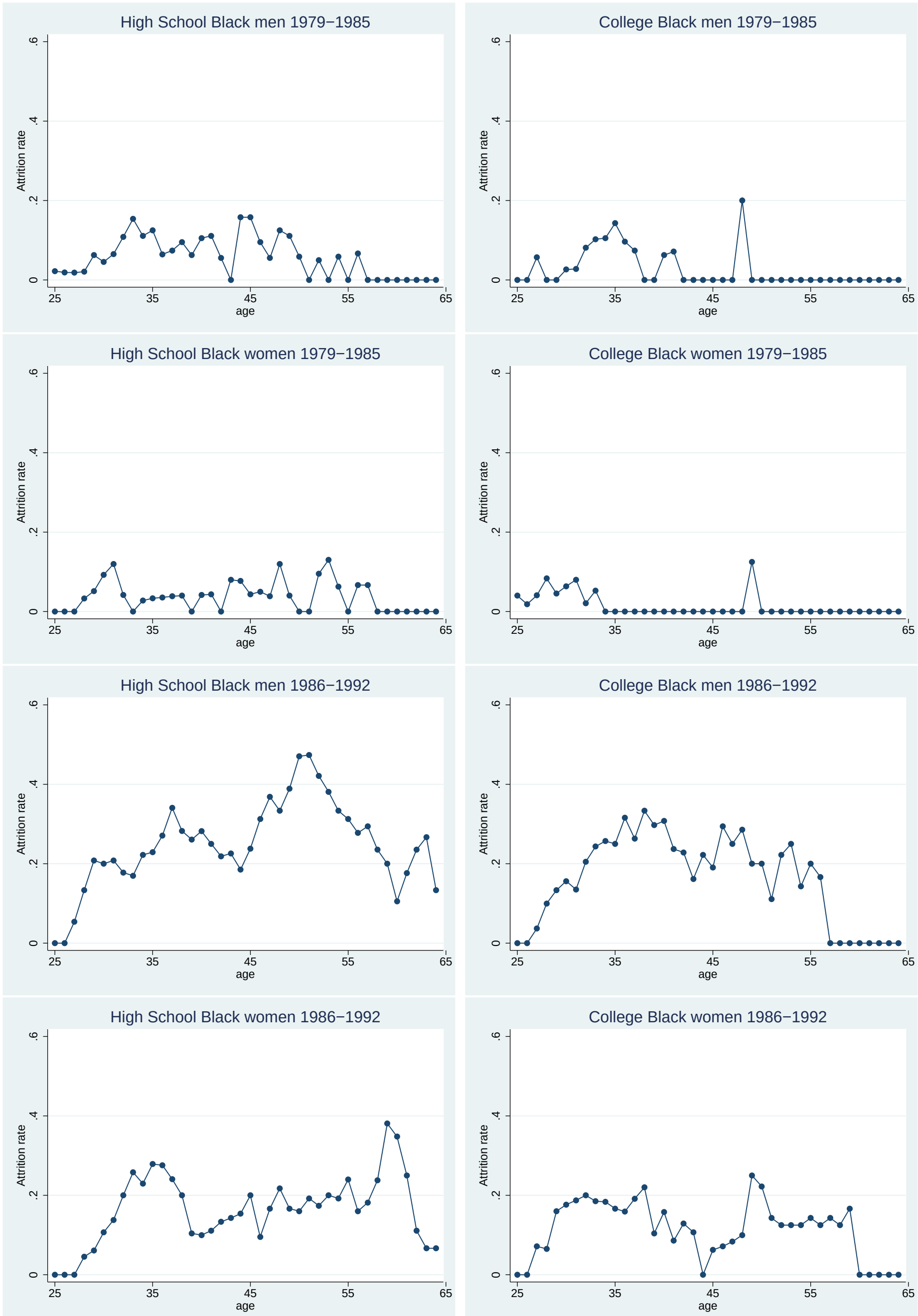


Figure 6: Attrition rates for Blacks %

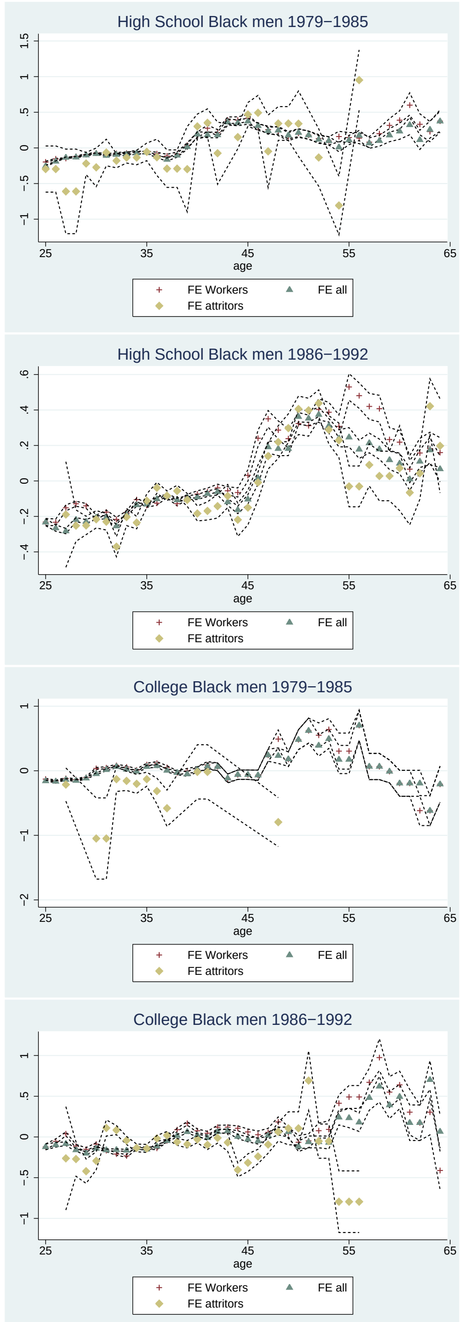


Figure 7: How different are attritors from those who stay: Fixed effects of Black men

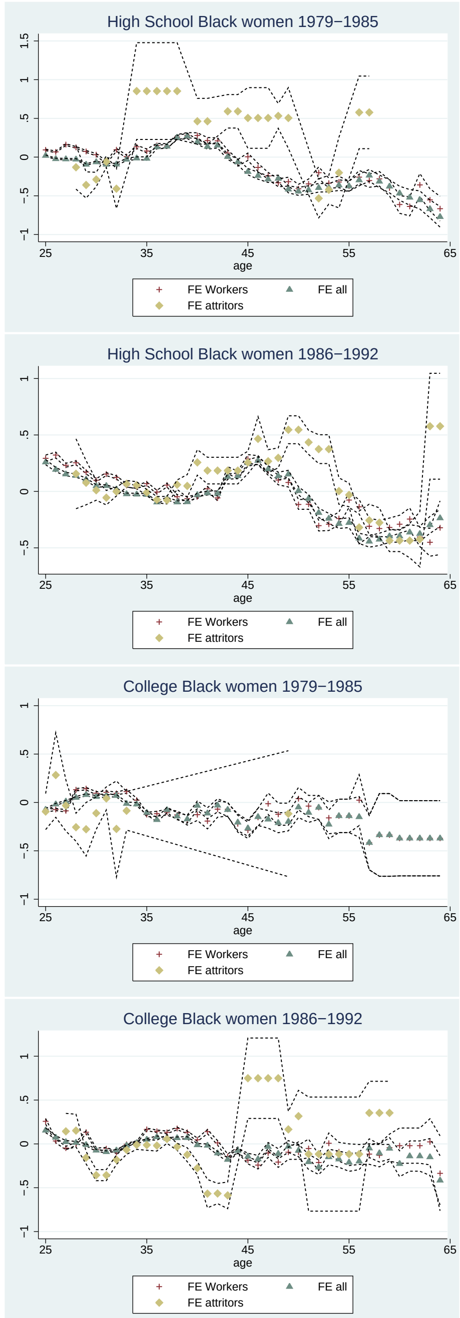


Figure 8: How different are attritors from those who stay: Fixed effects of Black women

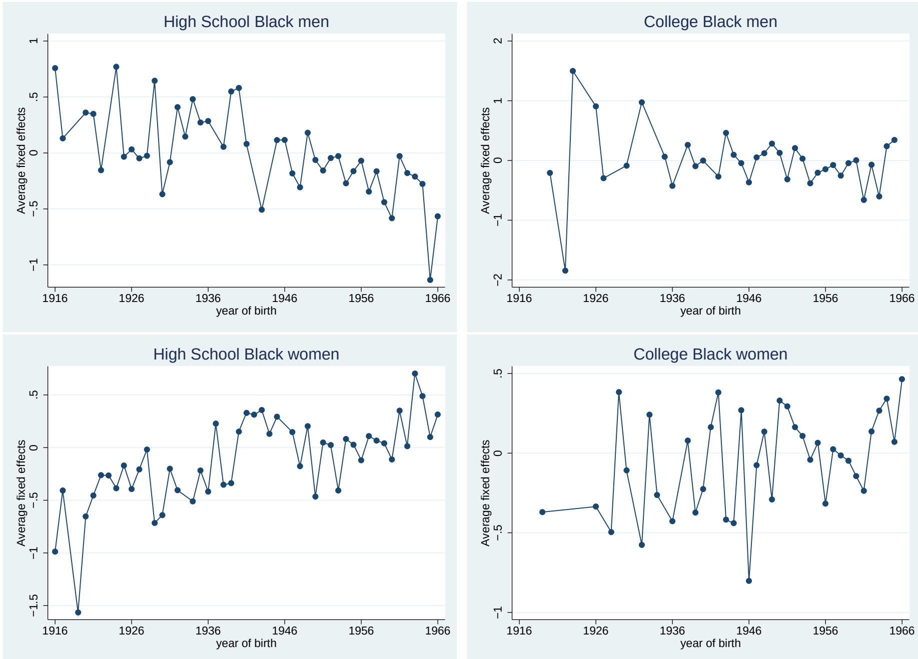


Figure 9: Average fixed effects by cohort 1919-1966

Appendix C

Estimated profiles correcting for attrition and cohort effects

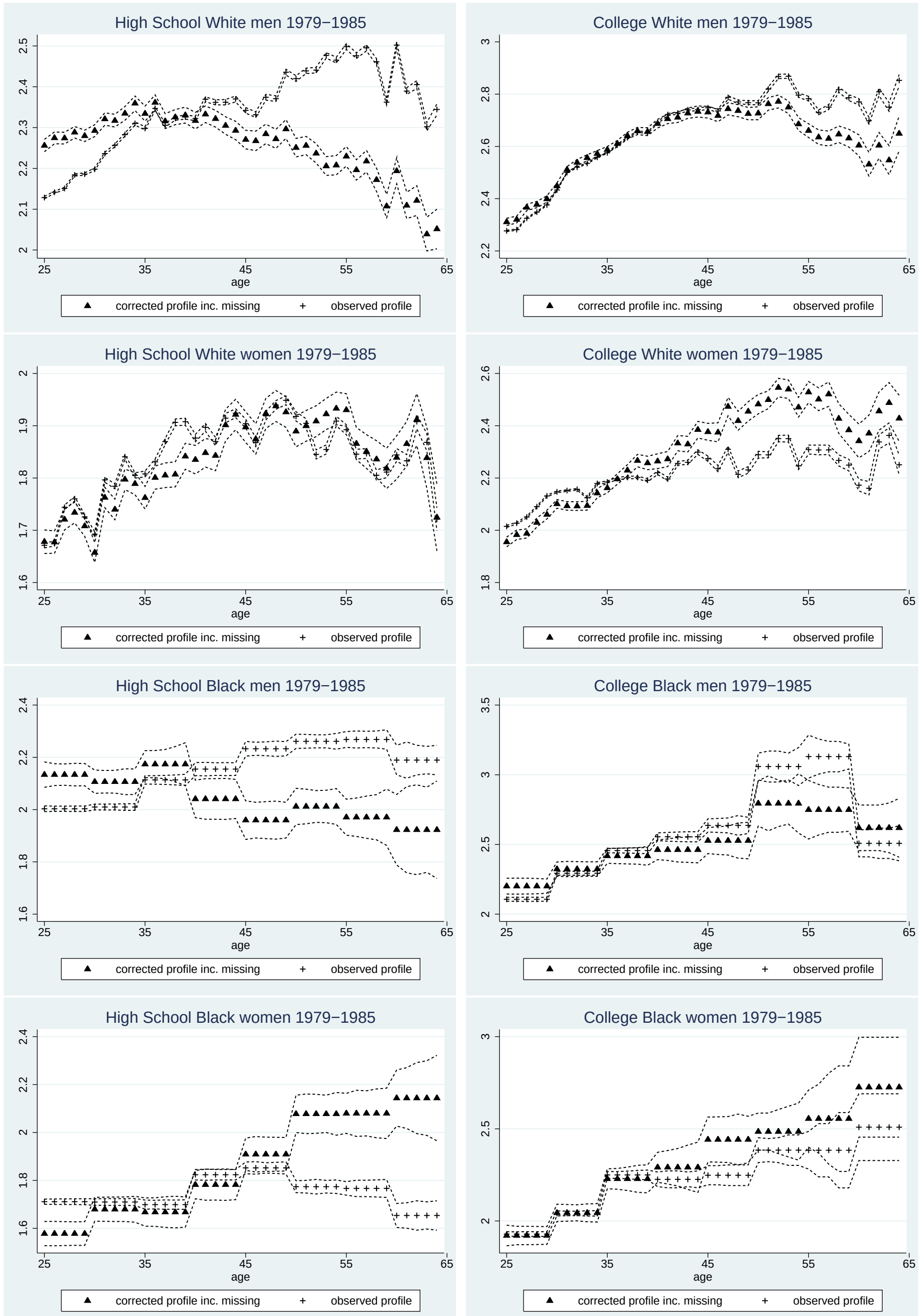


Figure 10: Accounting for attrition: Age-earnings profiles 1979-1985 using missing observations

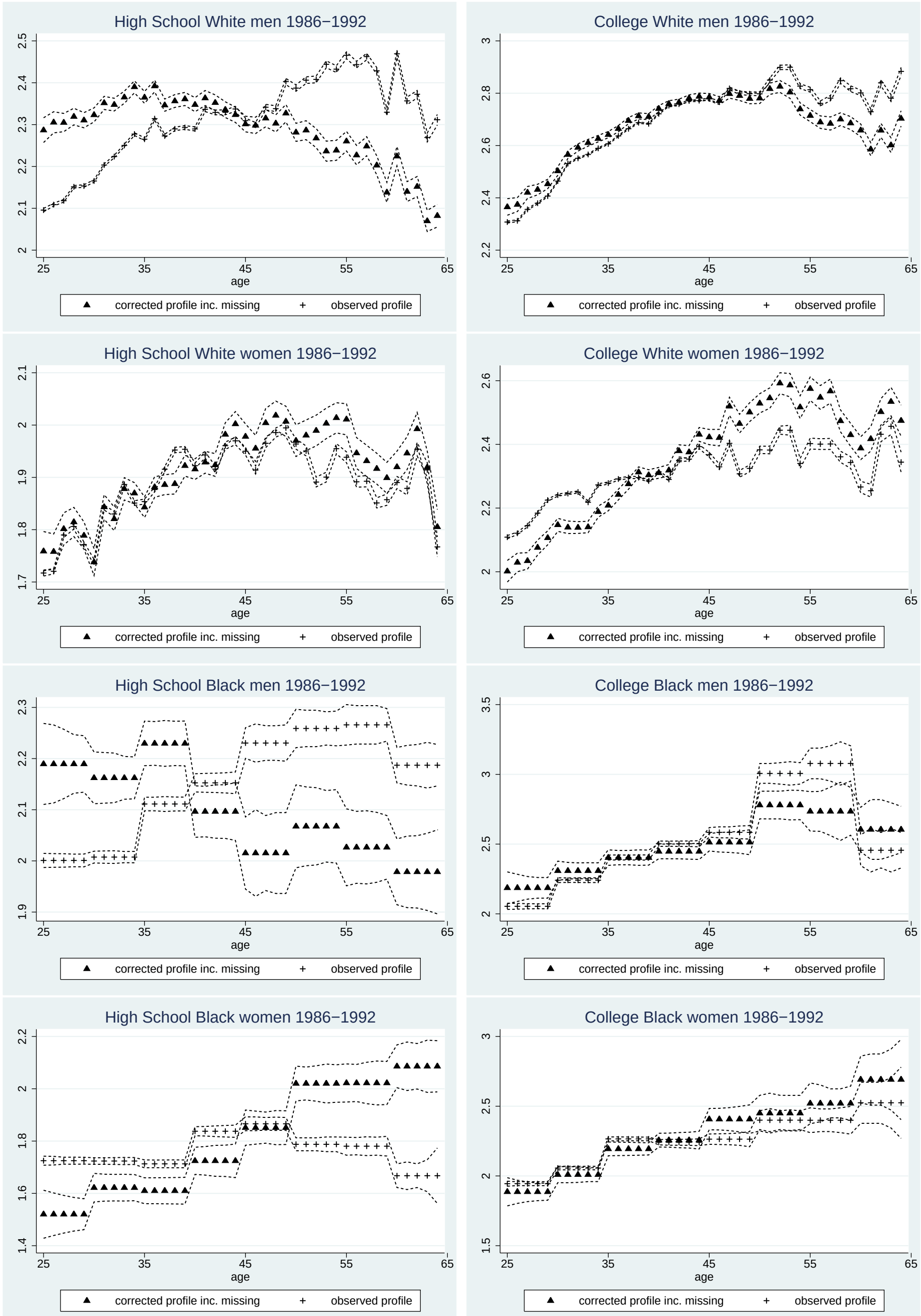


Figure 11: Accounting for attrition: Age-earnings profiles 1986-1992 using missing observations

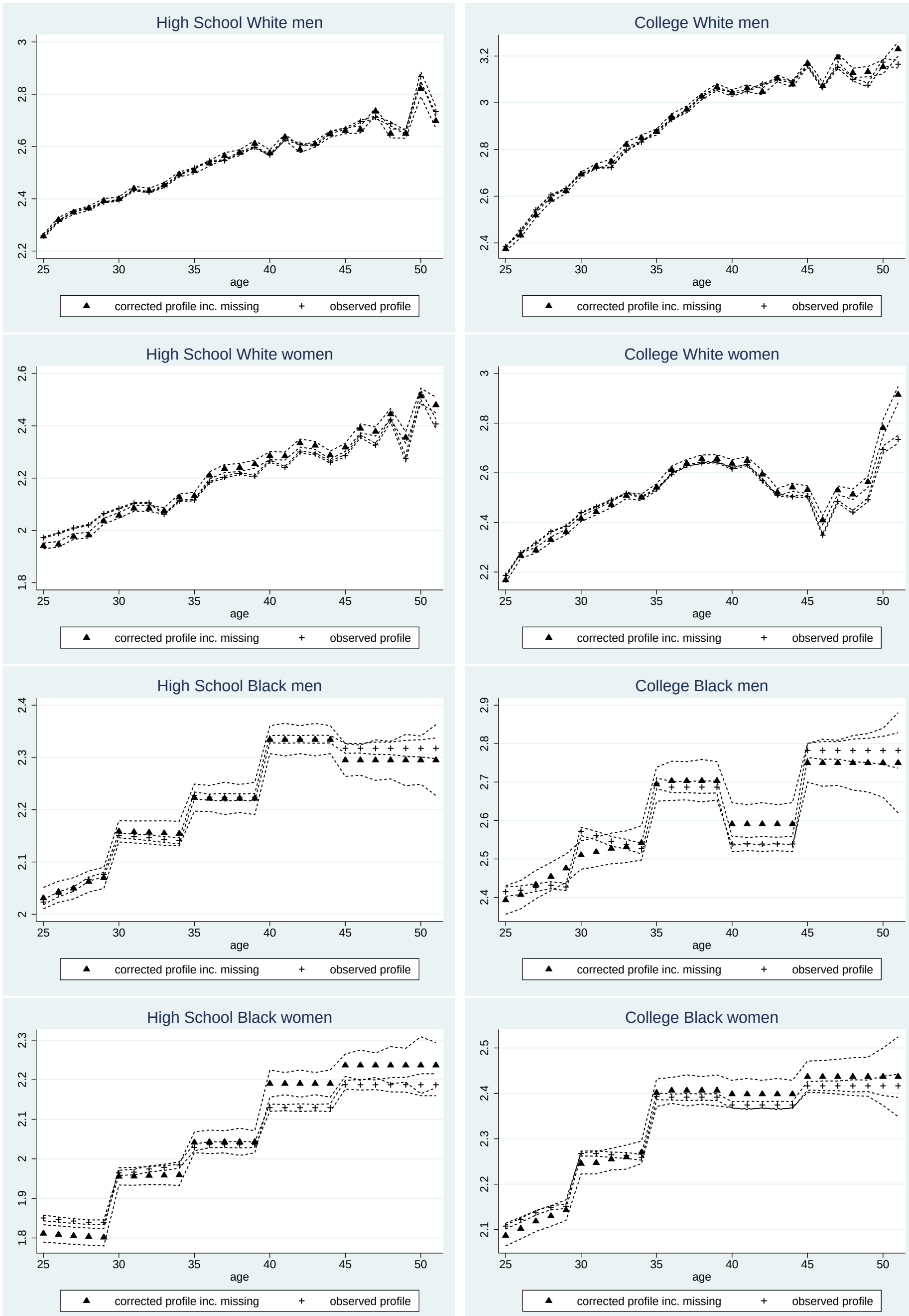


Figure 12: Accounting for attrition and cohort effects: Age-earnings profiles using NLSY and including missing observations

Appendix D

Intertemporal models with flexible labour supply

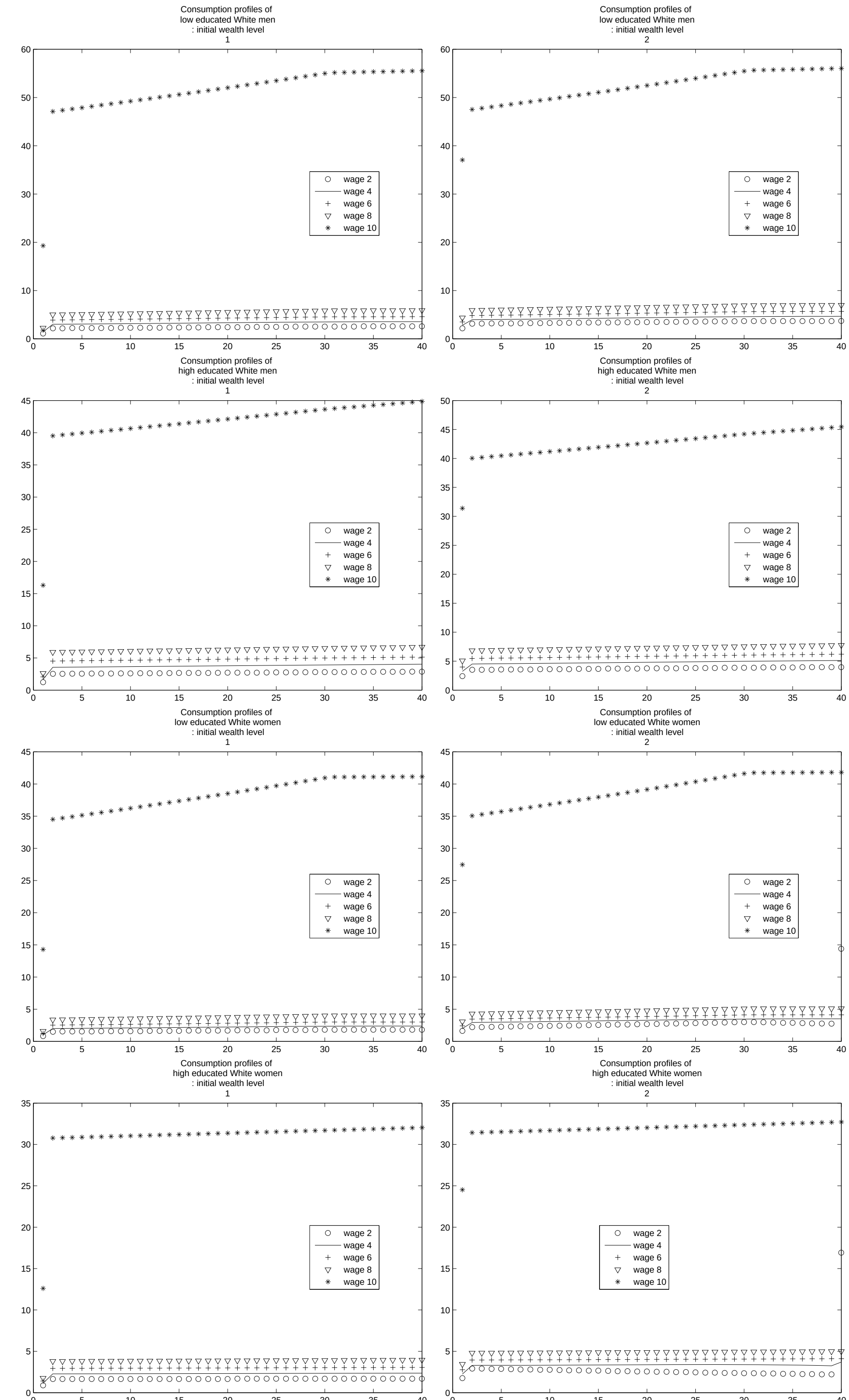
Section 1.5 showed the participation rates, hours of work and wage profiles simulated by three different lifecycle models with flexible labour supply. Here we show and discuss the detailed results of these simulations.

The basic intertemporal model with flexible labour supply

The following figures show the consumption, wealth and leisure profiles simulated for individuals with different productivity, conditional on wealth, for the basic model where wages are both exogenous and deterministic. Profiles of all eight groups are presented. For most groups, work tracks the wage schedule, as predicted in Heckman (1974a). Exception is to be made to the group with the highest wage levels, the College White men group, for whom leisure increases throughout the period, at an even higher rate once wages start decreasing, even though they always participate in the labour market. Those who do not participate in the labour market are the lowest earners, higher wealth individuals, typically the case of second earners as discussed in Neal (2004).

This tracking of wages is also observed in consumption. While wages are increasing, individual consumption is growing but after wages begin to fall, consumption growth slows down, stagnates or is reversed depending on the group. For the College White male group, consumption growth remains constant and for the group facing a wage constant negative wage growth rate (White College women), consumption decreases over the life course for low initial wage levels, but high enough wage levels allow for an initial consumption growth or even an increasing profile all along. This is achieved at the cost of large number of hours of work earlier in life and a sharp run down of wealth after an initial stock up. For the two groups facing very high wage growth rates (both High School Black groups), the cost of accumulating debt is so low that they it is used to fund steeply increasing consumption profiles earlier in life, and low labour market attachment.

In summary, the intertemporal substitution of leisure discussed in Heckman (1974a) is very strong in this model. Because individuals can change their labour supply hours, they choose to work longer hours when wages are higher. Those who have high wealth and low wages choose not to participate. Even though consumption and leisure are complements, these simulations showed consumption increasing when leisure is decreasing, i.e when wages are growing. This is probably due to low discounting. This model does not explain the decreasing participation observed in the data, unless the wage profile is decreasing, which is difficult to reconcile with human capital theories. It suggests discounting and having hours of work tracking wages are the two main drivers of participation decisions.


Consumption profiles of
low educated White men
: initial wealth level
1
Consumption profiles of
low educated White men
: initial wealth level
2
Consumption profiles of
high educated White men
: initial wealth level
1
Consumption profiles of
high educated White men
: initial wealth level
2
Consumption profiles of
low educated White women
: initial wealth level
1
Consumption profiles of
low educated White women
: initial wealth level
2
Consumption profiles of
high educated White women
: initial wealth level
1
Consumption profiles of
high educated White women
: initial wealth level
2
wage 2
wage 4
wage 6
wage 8
wage 10
0
5
10
15
20
25
30
35
40

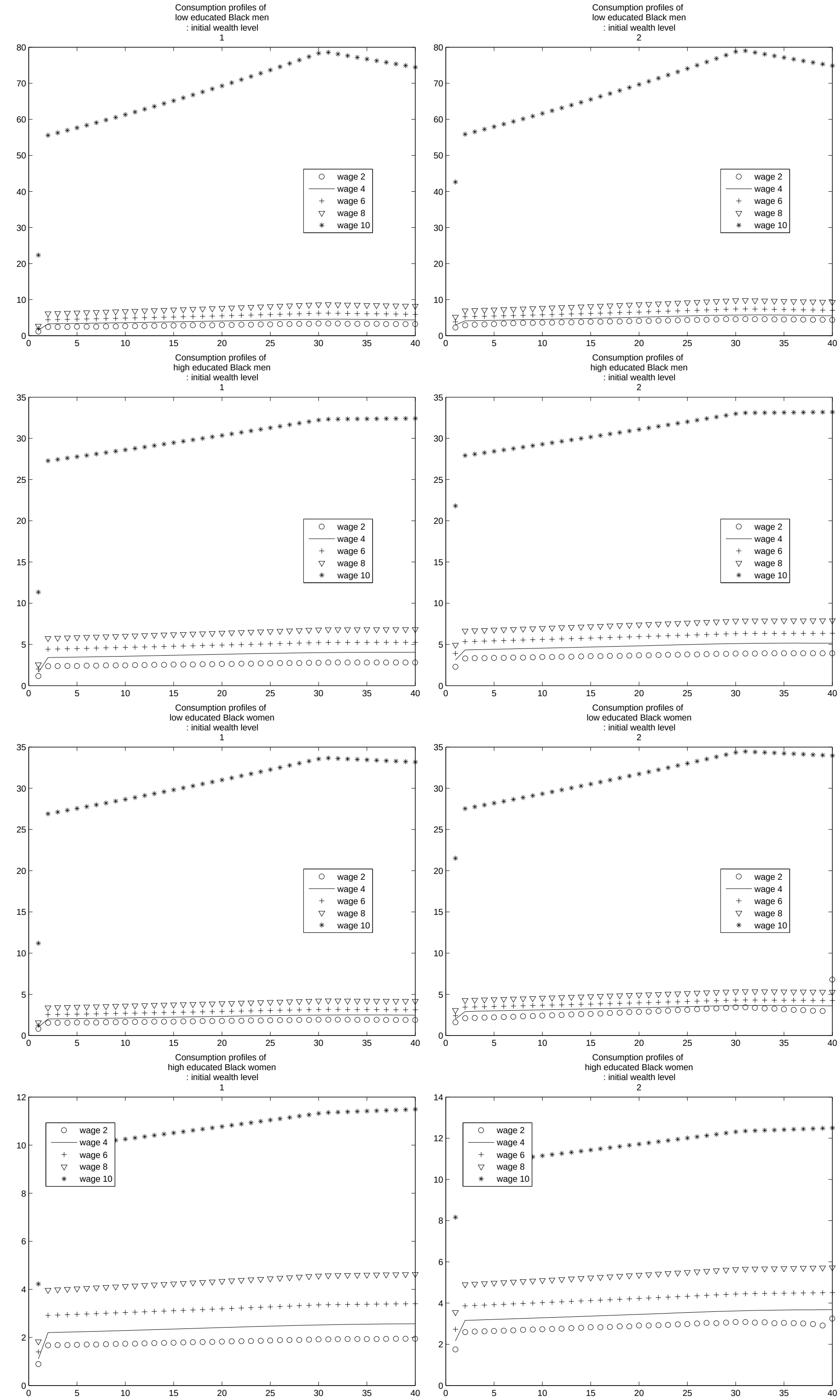
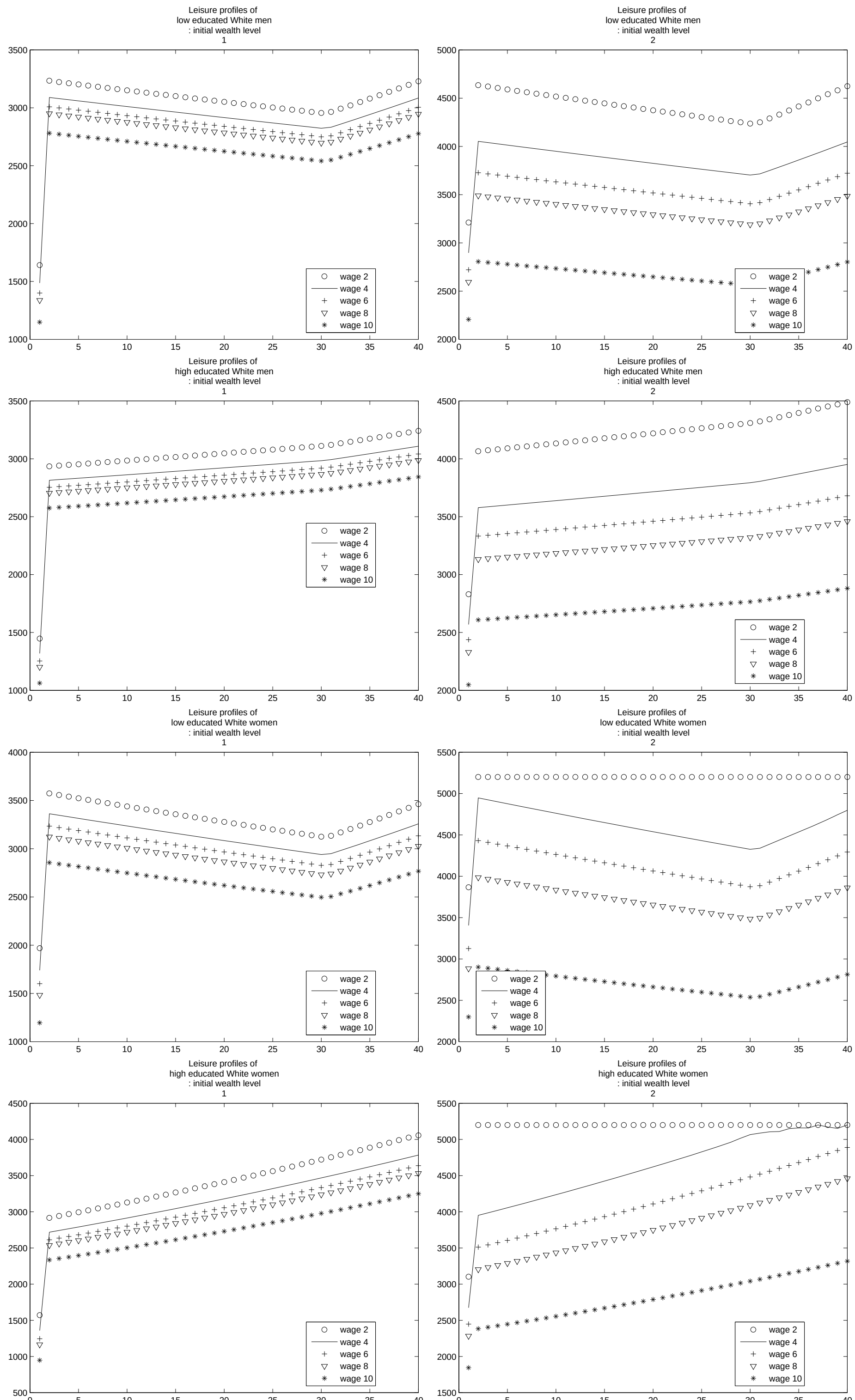


Figure 13: Consumption profiles in the basic model


Leisure profiles of
low educated White men
: initial wealth level
1
Leisure profiles of
low educated White men
: initial wealth level
2
Leisure profiles of
high educated White men
: initial wealth level
1
Leisure profiles of
high educated White men
: initial wealth level
2
Leisure profiles of
low educated White women
: initial wealth level
1
Leisure profiles of
low educated White women
: initial wealth level
2
Leisure profiles of
high educated White women
: initial wealth level
1
Leisure profiles of
high educated White women
: initial wealth level
2
wage 2
wage 4
wage 6
wage 8
wage 10

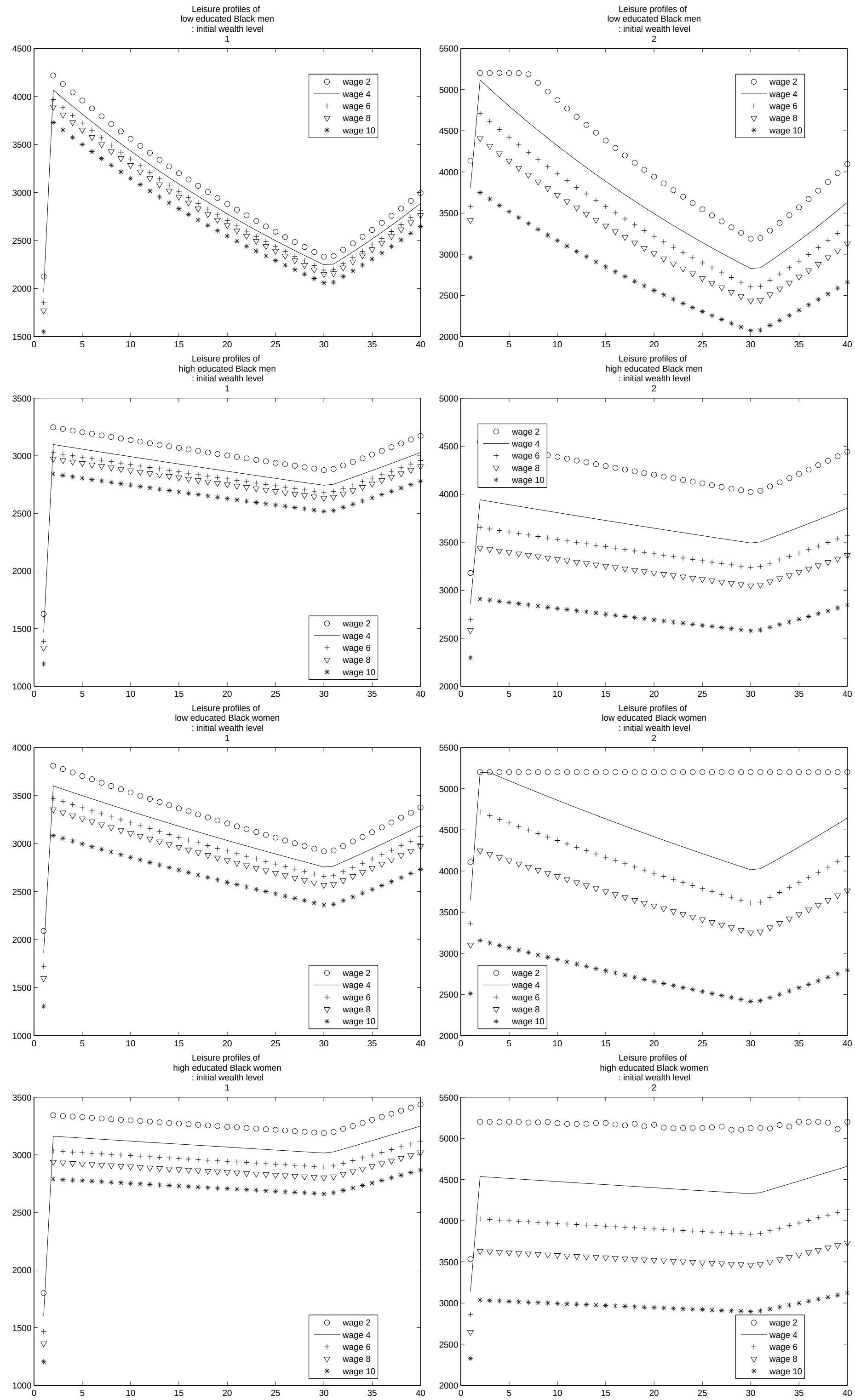
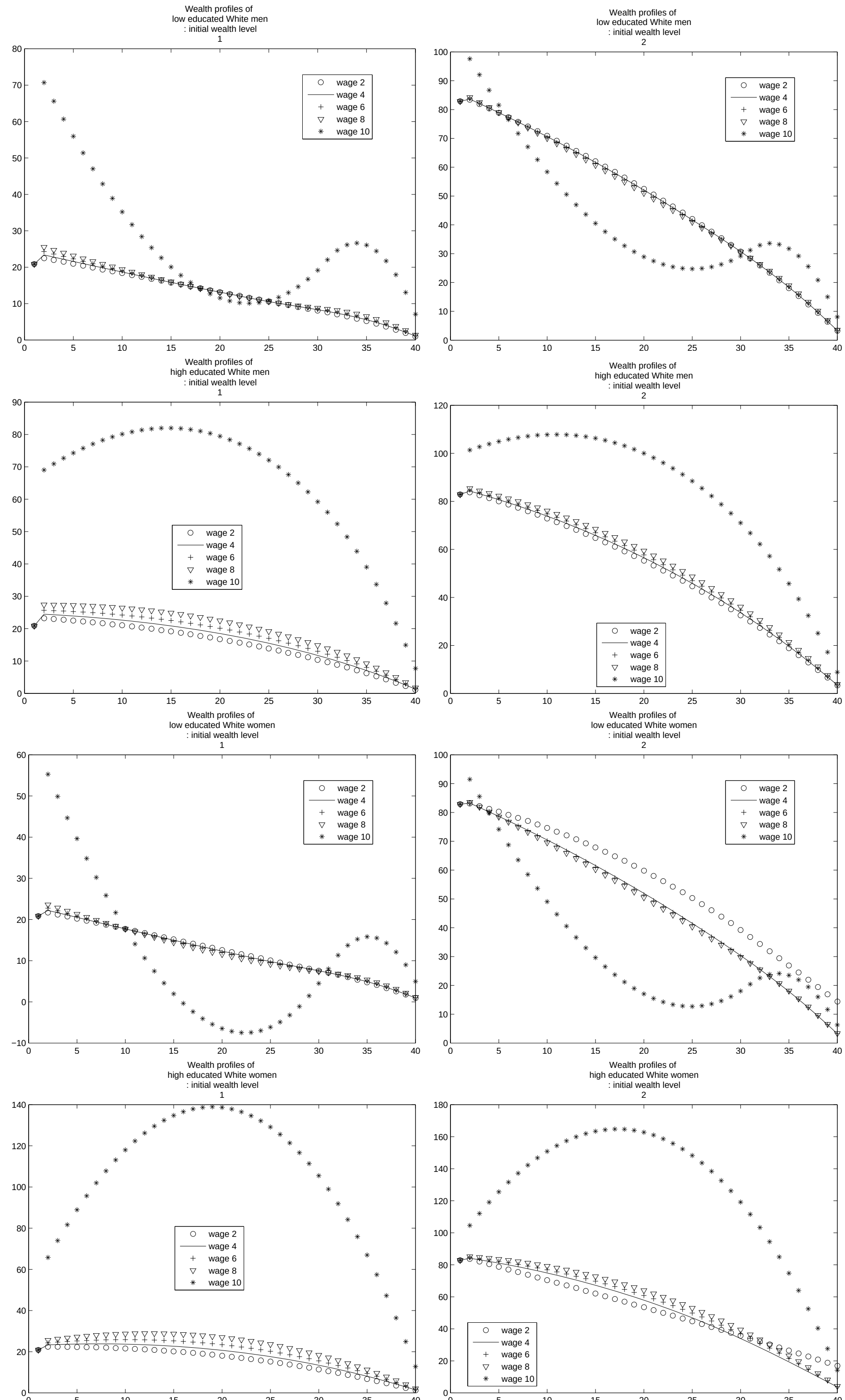


Figure 14: Leisure profiles in the basic model


Wealth profiles of
low educated White men
: initial wealth level
1
Wealth profiles of
low educated White men
: initial wealth level
2
Wealth profiles of
high educated White men
: initial wealth level
1
Wealth profiles of
high educated White men
: initial wealth level
2
Wealth profiles of
low educated White women
: initial wealth level
1
Wealth profiles of
low educated White women
: initial wealth level
2
Wealth profiles of
high educated White women
: initial wealth level
1
Wealth profiles of
high educated White women
: initial wealth level
2
wage 2
wage 4
wage 6
wage 8
wage 10

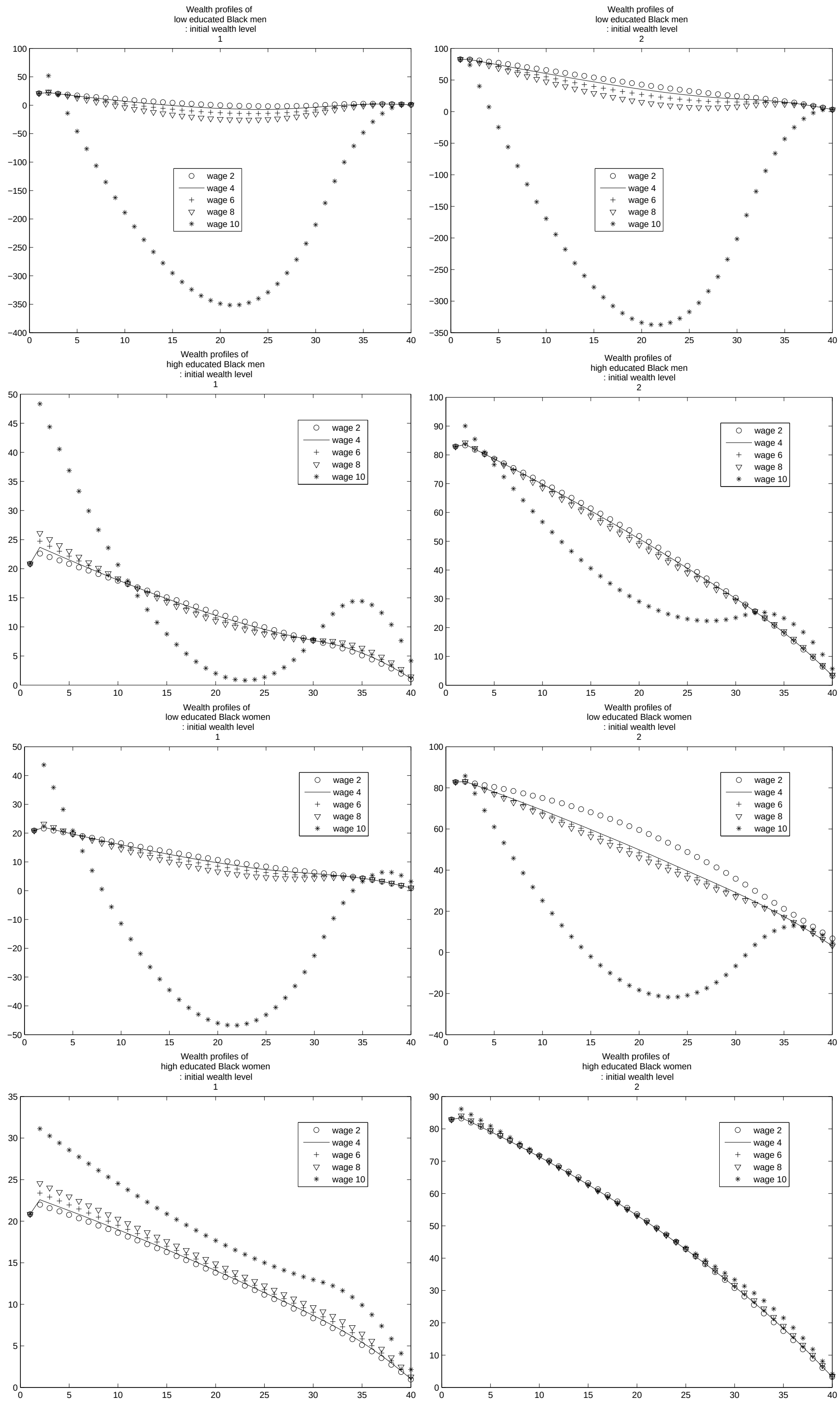
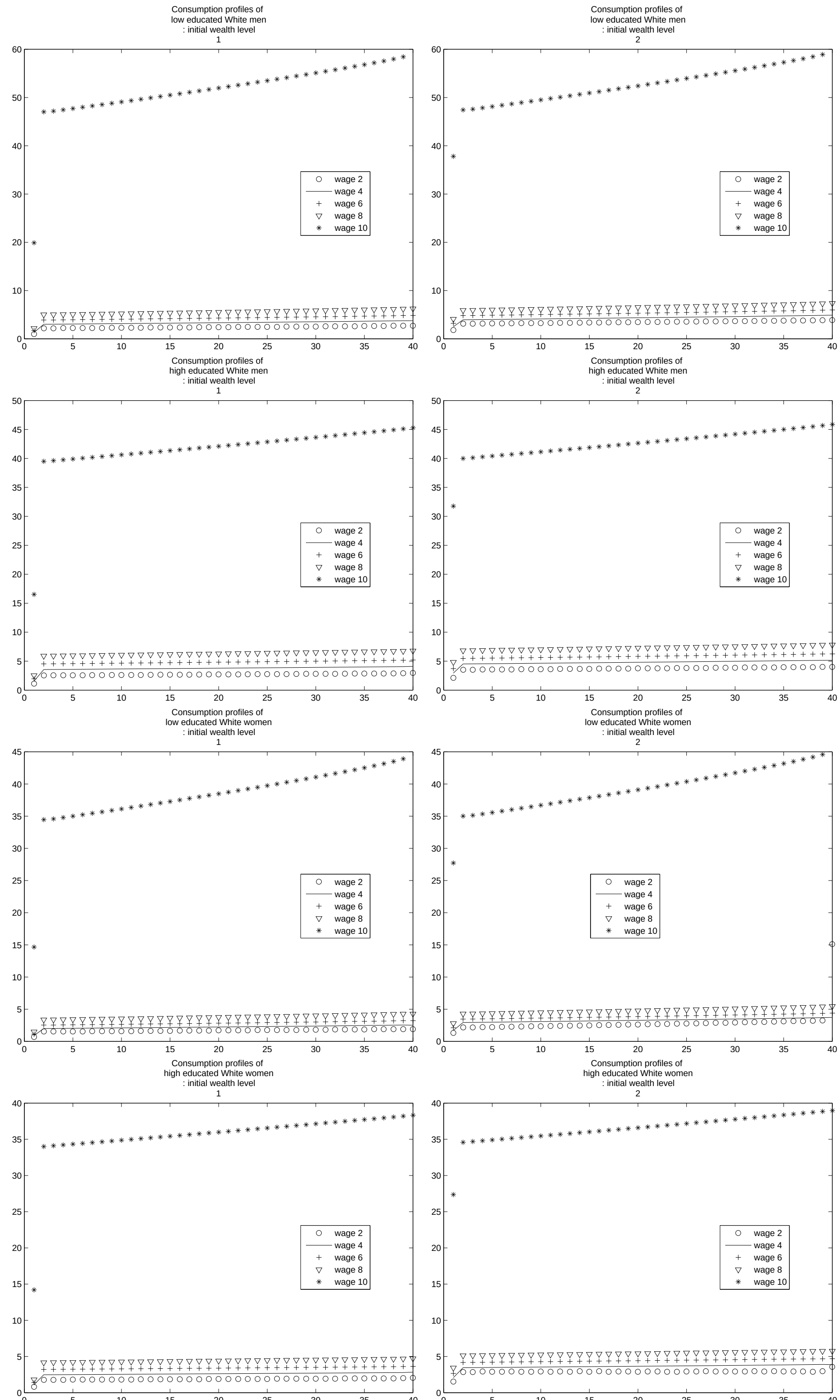


Figure 15: Wealth profiles in the basic model

The intertemporal model with flexible labour supply and deterministic endogenous wages

The following figures show the consumption, wealth and leisure profiles simulated for individuals with different productivity, conditional on wealth, for the model where wages are endogenous and deterministic. Profiles of all eight groups are presented. Consumption increases throughout the life course for all groups, except for low wage College White women who face no returns to experience in this simulation (their profiles are flat). These increasing profiles are financed differently by different groups. Groups facing a reasonably high wage growth rate will incur debt to finance consumption, while White College men and women will give up on leisure earlier in life.


Consumption profiles of
low educated White men
: initial wealth level
1
Consumption profiles of
low educated White men
: initial wealth level
2
Consumption profiles of
high educated White men
: initial wealth level
1
Consumption profiles of
high educated White men
: initial wealth level
2
Consumption profiles of
low educated White women
: initial wealth level
1
Consumption profiles of
low educated White women
: initial wealth level
2
Consumption profiles of
high educated White women
: initial wealth level
1
Consumption profiles of
high educated White women
: initial wealth level
2
wage 2
wage 4
wage 6
wage 8
wage 10

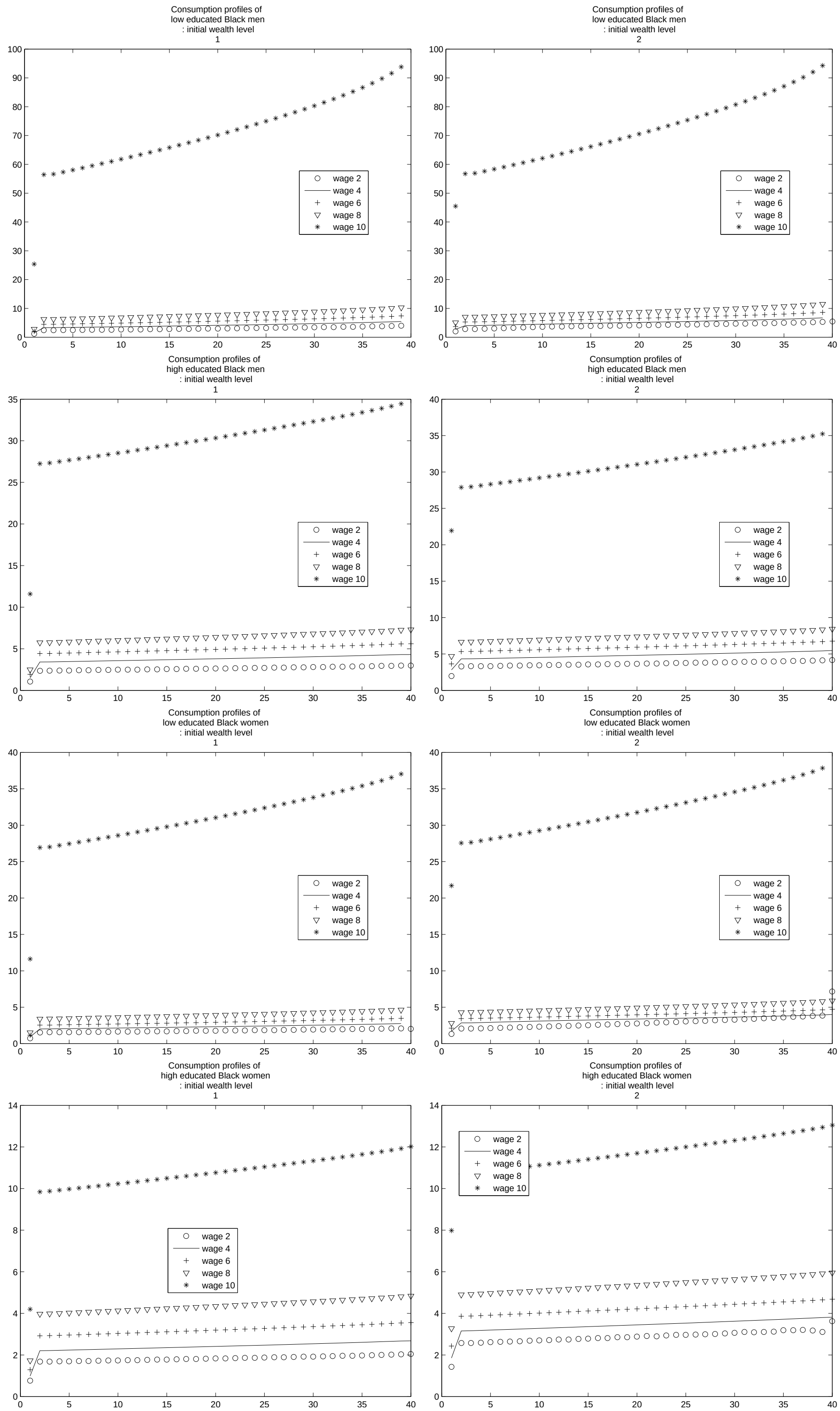
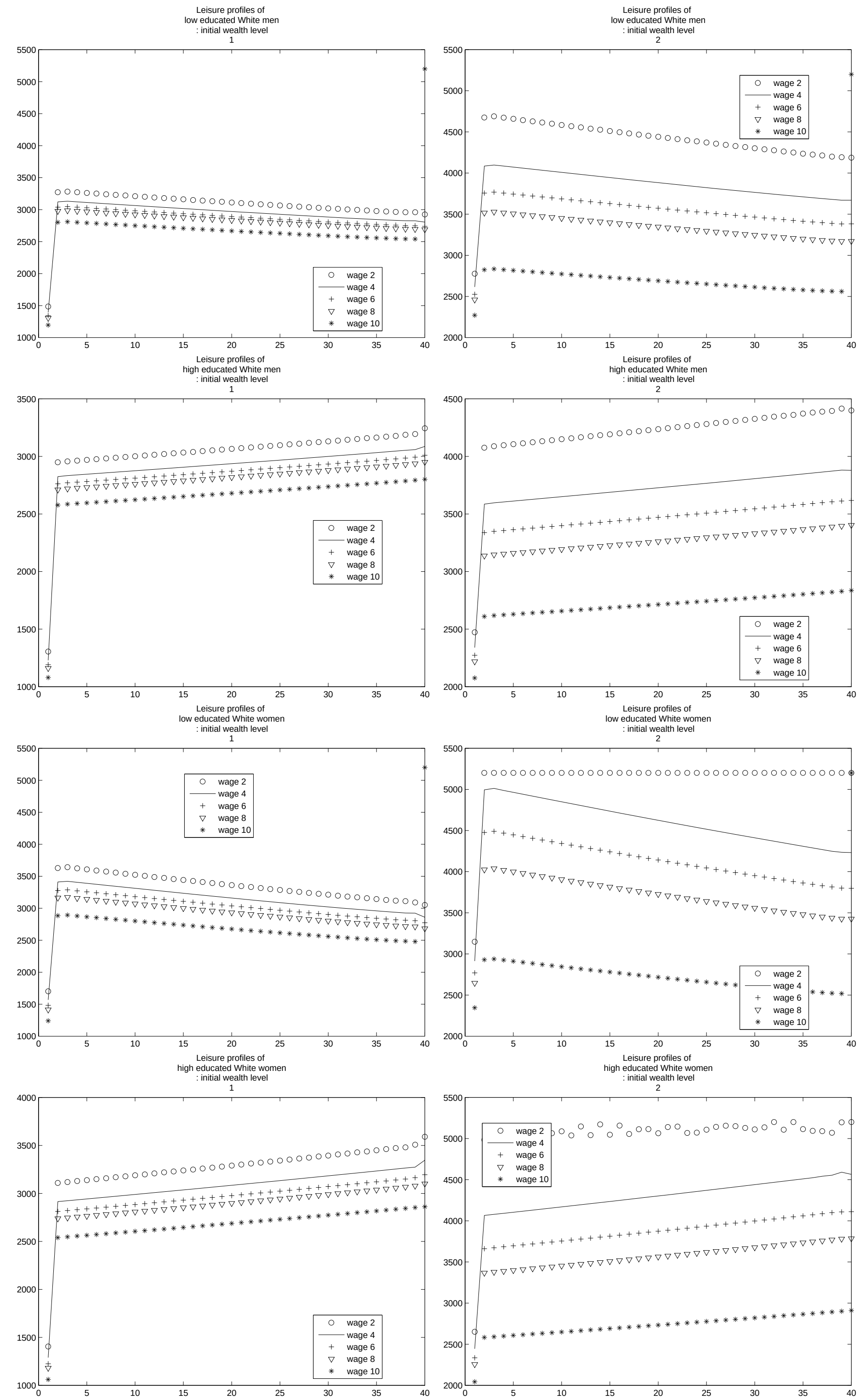


Figure 16: Consumption profiles in the model with deterministic endogenous wages



Leisure profiles of
low educated White men
: initial wealth level
1
Leisure profiles of
low educated White men
: initial wealth level
2
Leisure profiles of
high educated White men
: initial wealth level
1
Leisure profiles of
high educated White men
: initial wealth level
2
Leisure profiles of
low educated White women
: initial wealth level
1
Leisure profiles of
low educated White women
: initial wealth level
2
Leisure profiles of
high educated White women
: initial wealth level
1
Leisure profiles of
high educated White women
: initial wealth level
2
wage 2
wage 4
wage 6
wage 8
wage 10

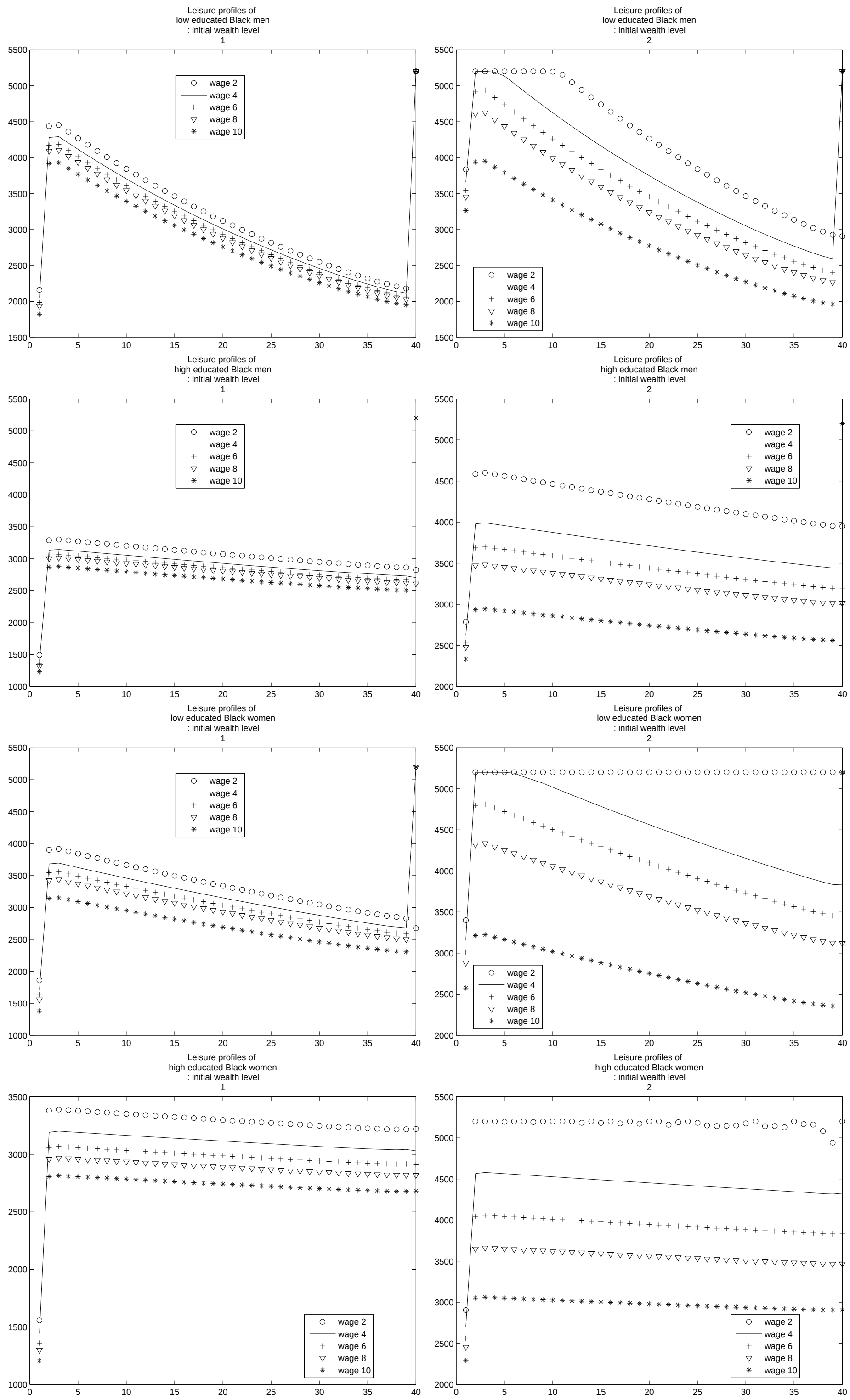
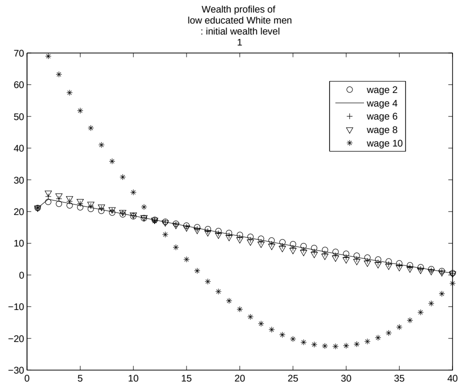
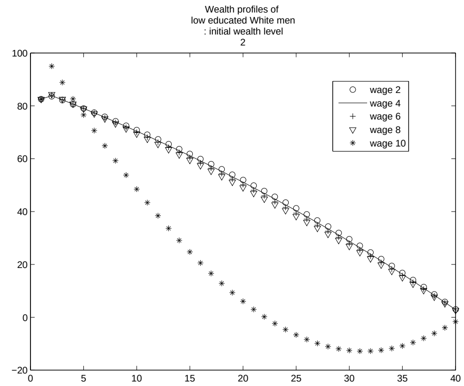
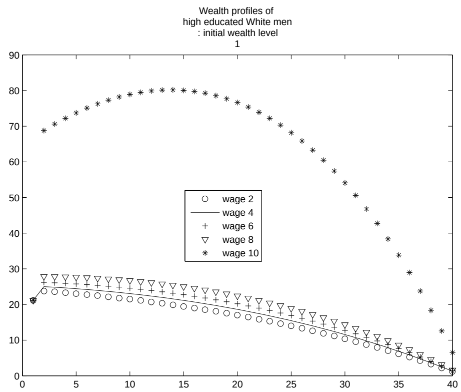
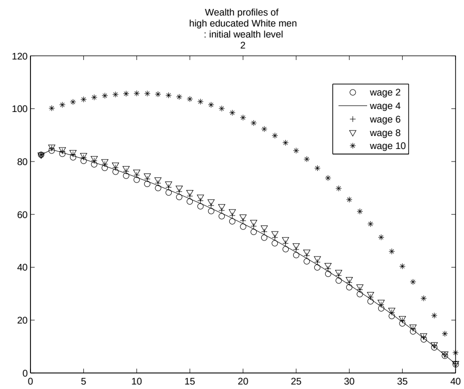


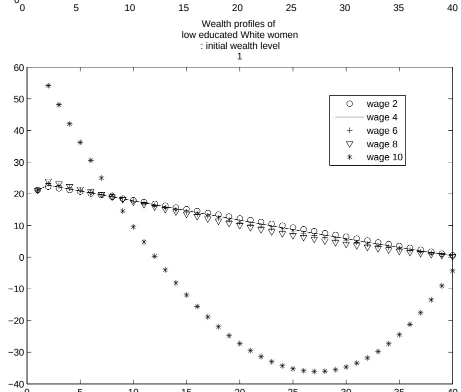
Figure 17: Leisure profiles in the model with deterministic endogenous wages

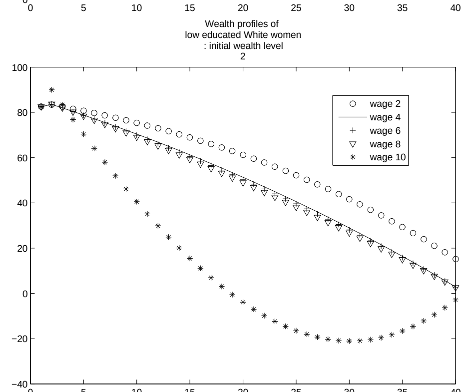

Wealth profiles of
low educated White men
: initial wealth level
1
wage 2
wage 4
wage 6
wage 8
wage 10

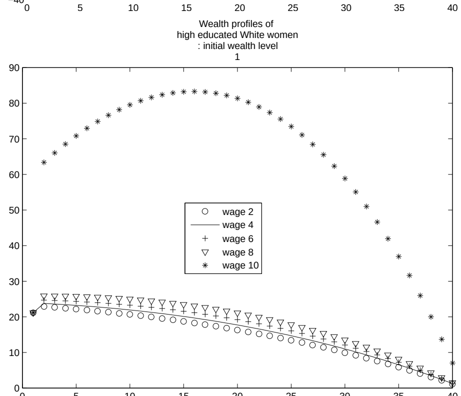

Wealth profiles of
low educated White men
: initial wealth level
2
wage 2
wage 4
wage 6
wage 8
wage 10

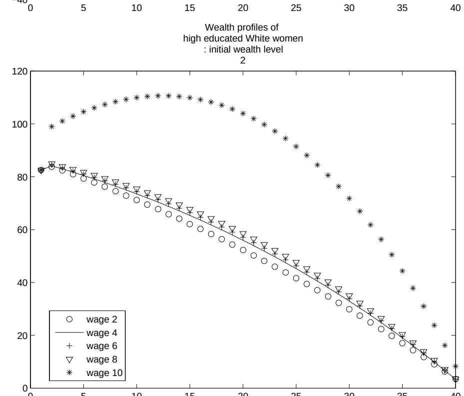

Wealth profiles of
high educated White men
: initial wealth level
1
wage 2
wage 4
wage 6
wage 8
wage 10


Wealth profiles of
high educated White men
: initial wealth level
2
wage 2
wage 4
wage 6
wage 8
wage 10


Wealth profiles of
low educated White women
: initial wealth level
1
wage 2
wage 4
wage 6
wage 8
wage 10


Wealth profiles of
low educated White women
: initial wealth level
2
wage 2
wage 4
wage 6
wage 8
wage 10


Wealth profiles of
high educated White women
: initial wealth level
1
wage 2
wage 4
wage 6
wage 8
wage 10


Wealth profiles of
high educated White women
: initial wealth level
2
wage 2
wage 4
wage 6
wage 8
wage 10

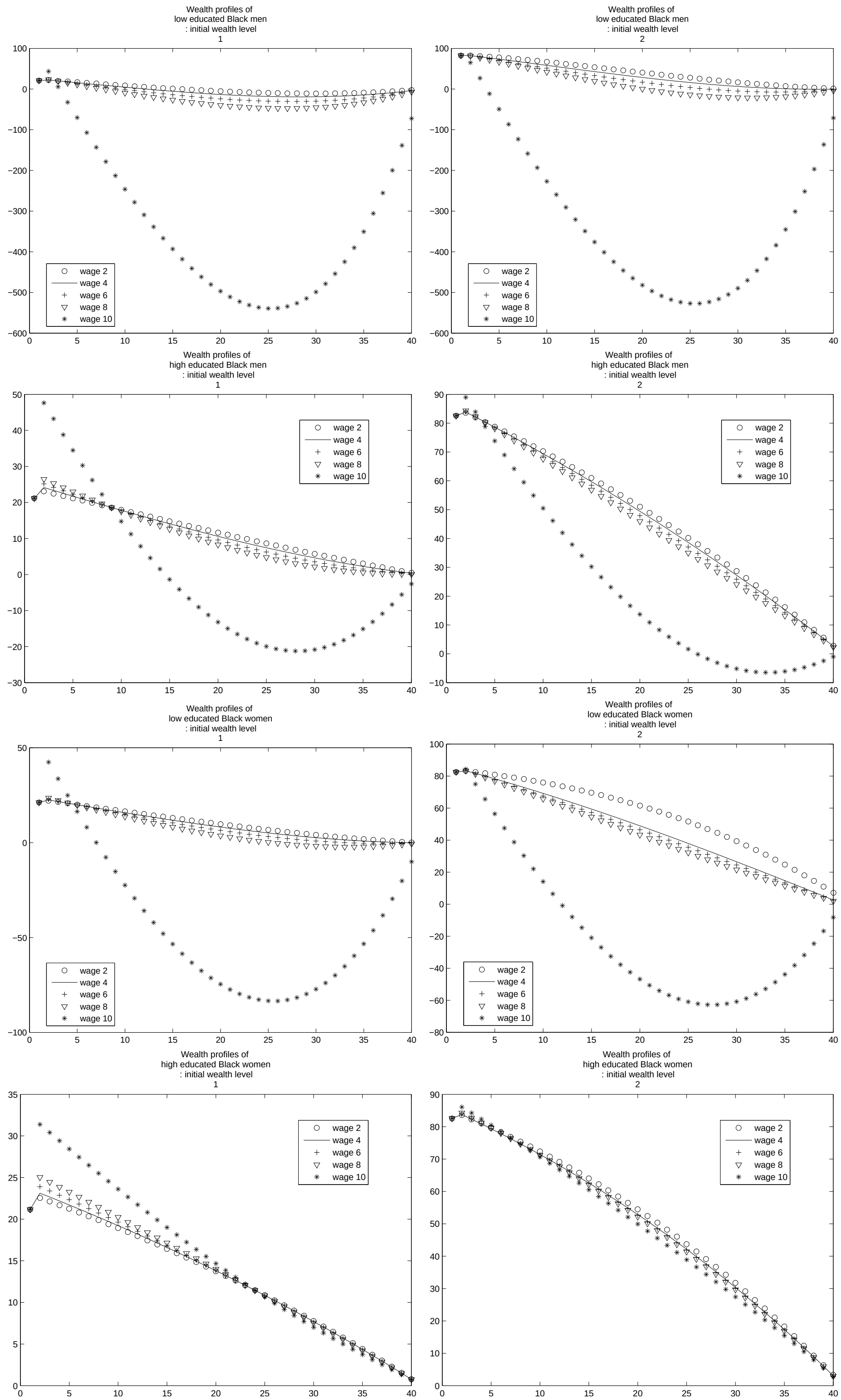
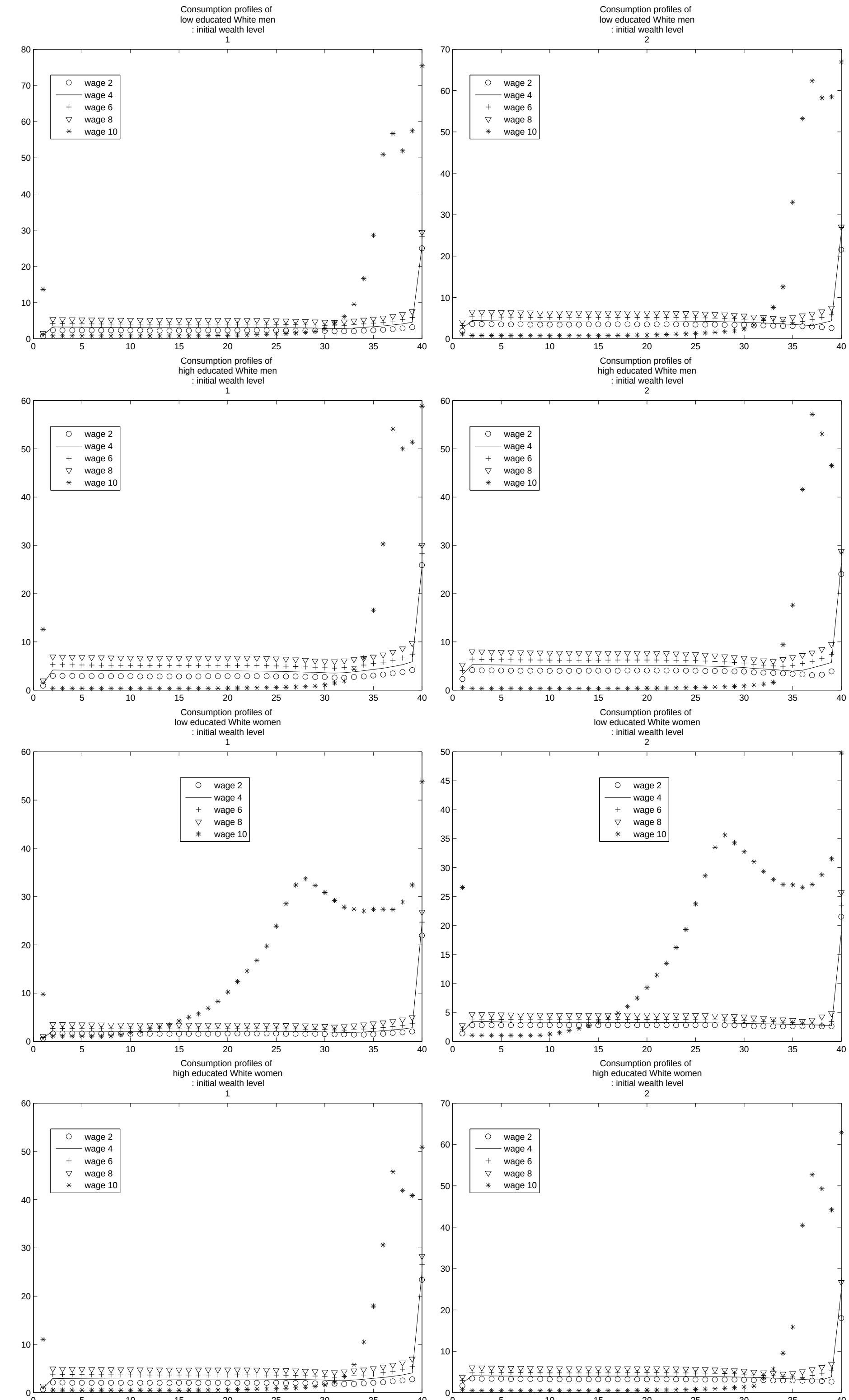


Figure 18: Wealth profiles in the model with deterministic endogenous wages

The intertemporal model with flexible labour supply and stochastic exogenous wages

The following figures show the consumption, wealth and leisure profiles simulated for individuals with different productivity, conditional on wealth, for the model where wages are exogenous and stochastic. Profiles of all eight groups are presented. We generated a sequence of 40 random shocks to wages for each of 5000 individuals within each group. At the end of each period, a shock to the wage occurs, which translates into a shock to the wealth stock available at the beginning of the next period. The profiles represent the average optimal consumption and leisure choices made by these 5000 individuals who experience different wage trajectories, and the wealth resulting from these choices. Consumption profiles are fairly flat, almost always decreasing until wage shocks become riskier. Then consumption suffers a drop and comes back up again at the end of life, when uncertainty is dissipated and the wealth buffer against shocks is no longer needed. For very high earners, consumption increases throughout the life course, but at a slower rate than wealth accumulates. These individuals, unless their high wage growth rates offset the risk of a negative shock, stock up huge wealth which peaks just before midlife. At the end of life, and because they stop working, wealth is run down fast. For those with modest wage levels or growth rates, wealth is run down from the very beginning, despite a slight attempt to buffer shocks. (Virtually) no debt is incurred. Leisure profiles are beyond doubt the most homogenous across groups. There is clear substitutability between leisure and consumption. Leisure was sacrificed in the beginning of life to allow wealth accumulation and insurance against adverse shocks, and increases very substantially over age for low initial wealth individuals. For high wealth individuals who can insure against shocks more easily, this pattern is not there.


Consumption profiles of
low educated White men
: initial wealth level
1
Consumption profiles of
low educated White men
: initial wealth level
2
Consumption profiles of
high educated White men
: initial wealth level
1
Consumption profiles of
high educated White men
: initial wealth level
2
Consumption profiles of
low educated White women
: initial wealth level
1
Consumption profiles of
low educated White women
: initial wealth level
2
Consumption profiles of
high educated White women
: initial wealth level
1
Consumption profiles of
high educated White women
: initial wealth level
2
wage 2
wage 4
wage 6
wage 8
wage 10

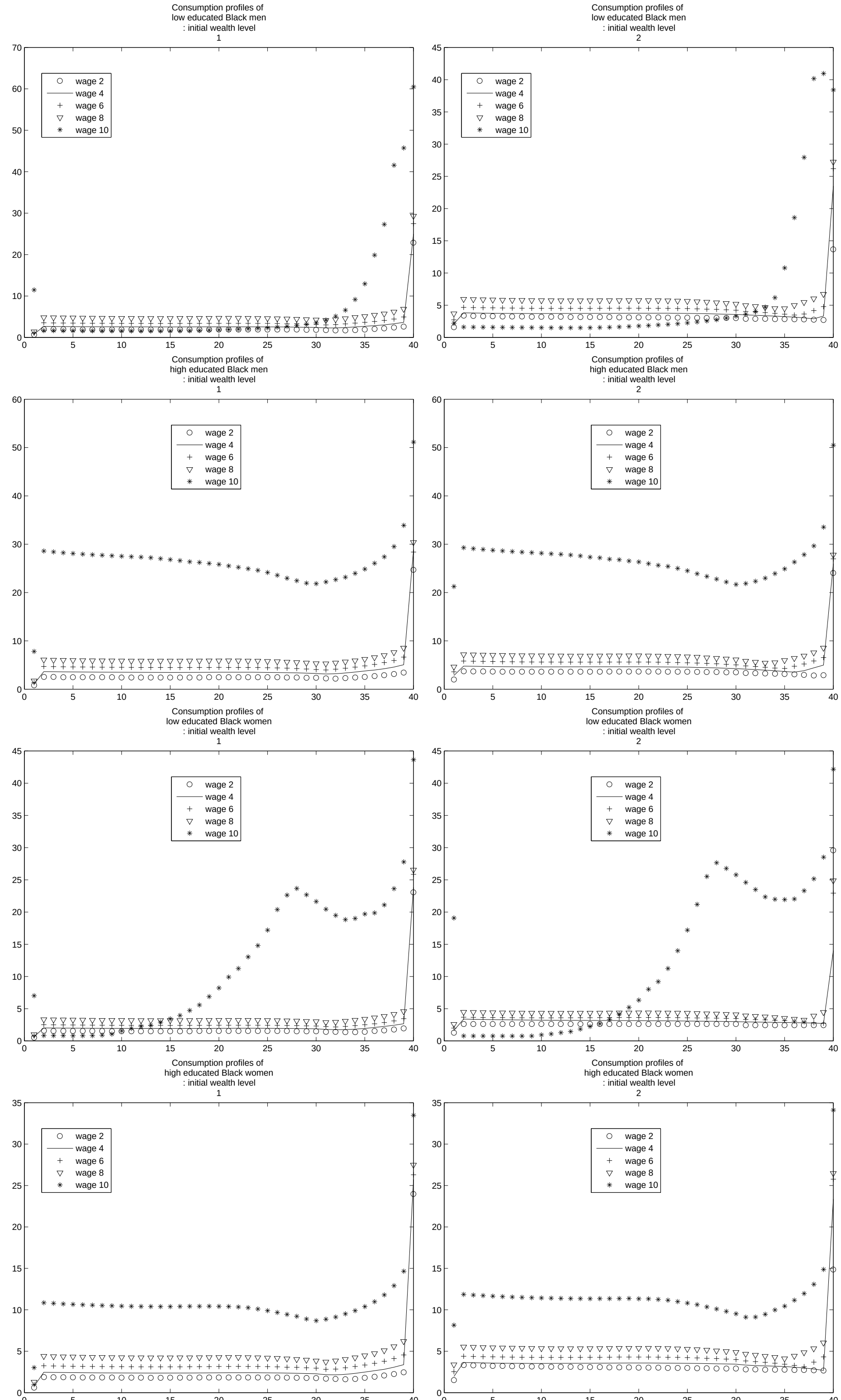
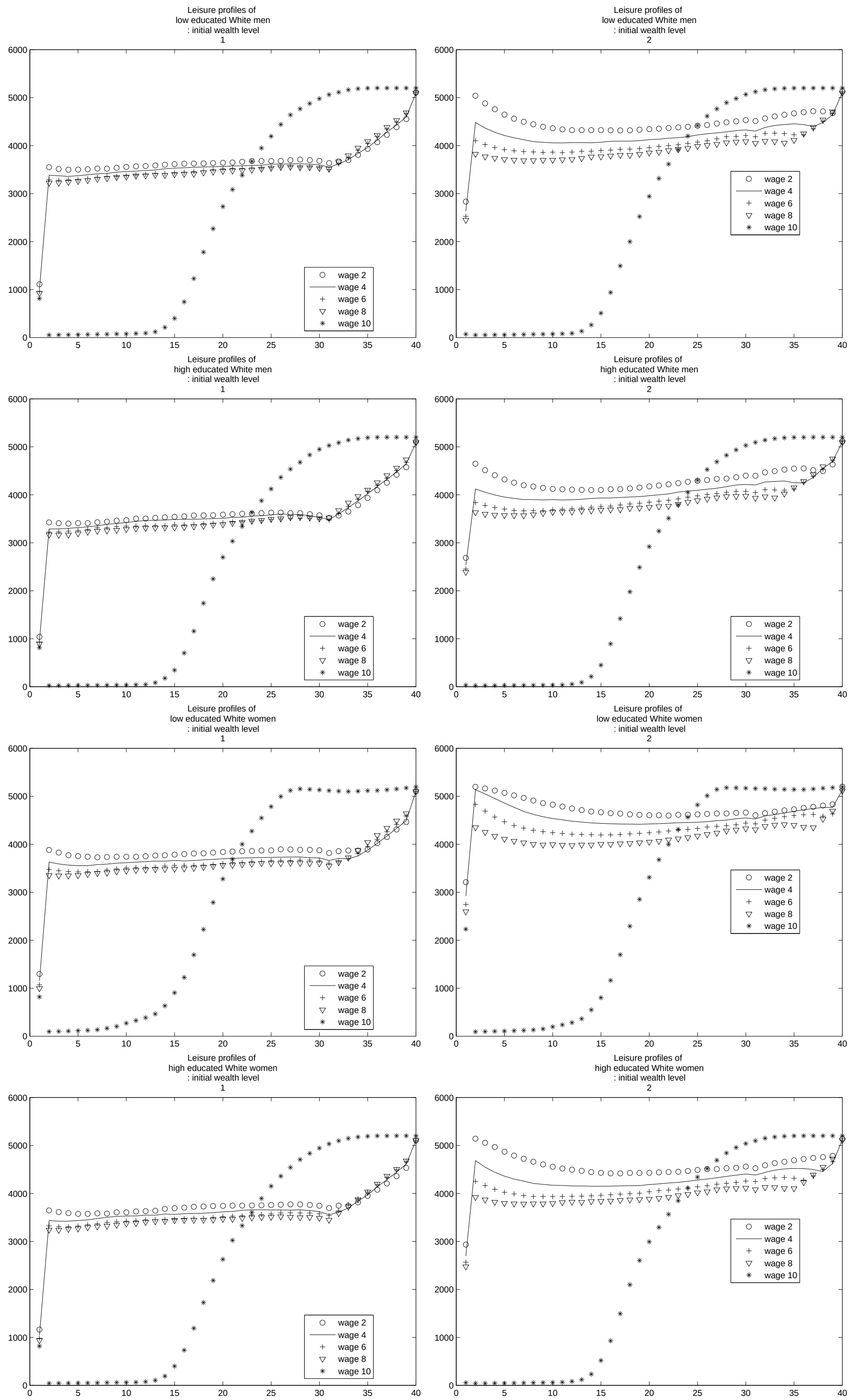


Figure 19: Consumption profiles in the model with uncertainty in wages


Leisure profiles of
low educated White men
: initial wealth level
1
Leisure profiles of
low educated White men
: initial wealth level
2
Leisure profiles of
high educated White men
: initial wealth level
1
Leisure profiles of
high educated White men
: initial wealth level
2
Leisure profiles of
low educated White women
: initial wealth level
1
Leisure profiles of
low educated White women
: initial wealth level
2
Leisure profiles of
high educated White women
: initial wealth level
1
Leisure profiles of
high educated White women
: initial wealth level
2
wage 2
wage 4
wage 6
wage 8
wage 10

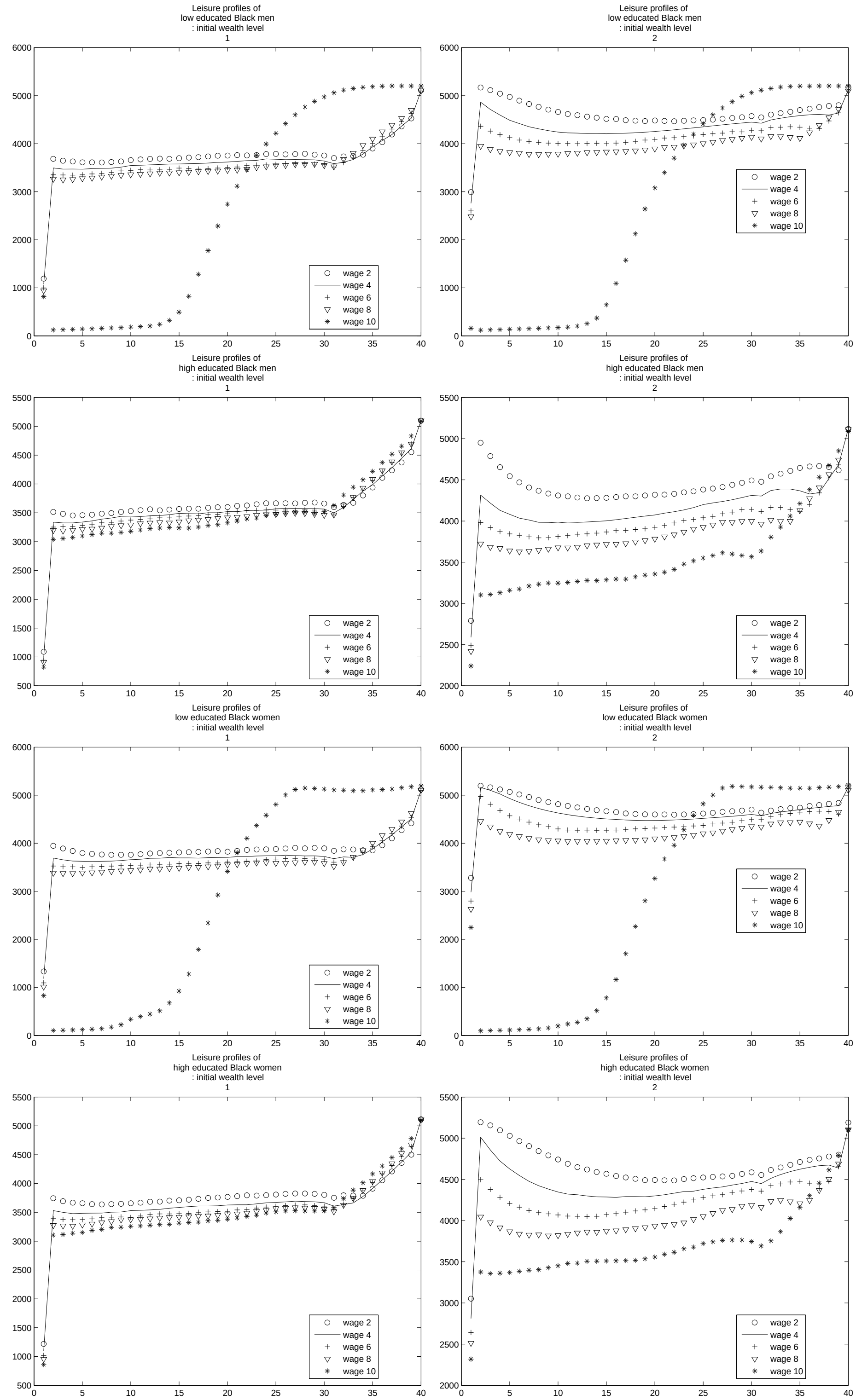
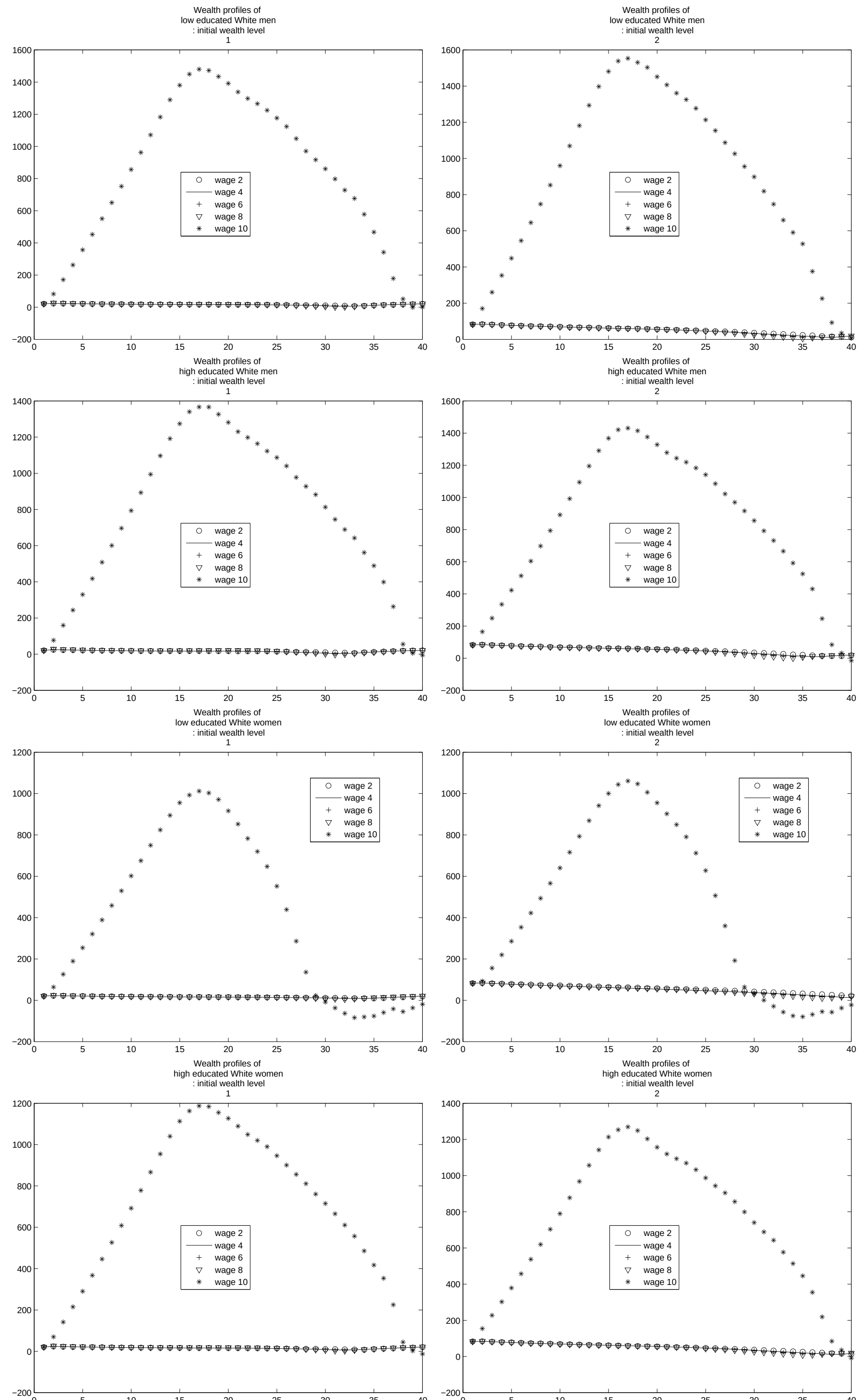


Figure 20: Leisure profiles in the model with uncertainty in wages


Wealth profiles of
low educated White men
: initial wealth level
1
Wealth profiles of
low educated White men
: initial wealth level
2
Wealth profiles of
high educated White men
: initial wealth level
1
Wealth profiles of
high educated White men
: initial wealth level
2
Wealth profiles of
low educated White women
: initial wealth level
1
Wealth profiles of
low educated White women
: initial wealth level
2
Wealth profiles of
high educated White women
: initial wealth level
1
Wealth profiles of
high educated White women
: initial wealth level
2
wage 2
wage 4
wage 6
wage 8
wage 10

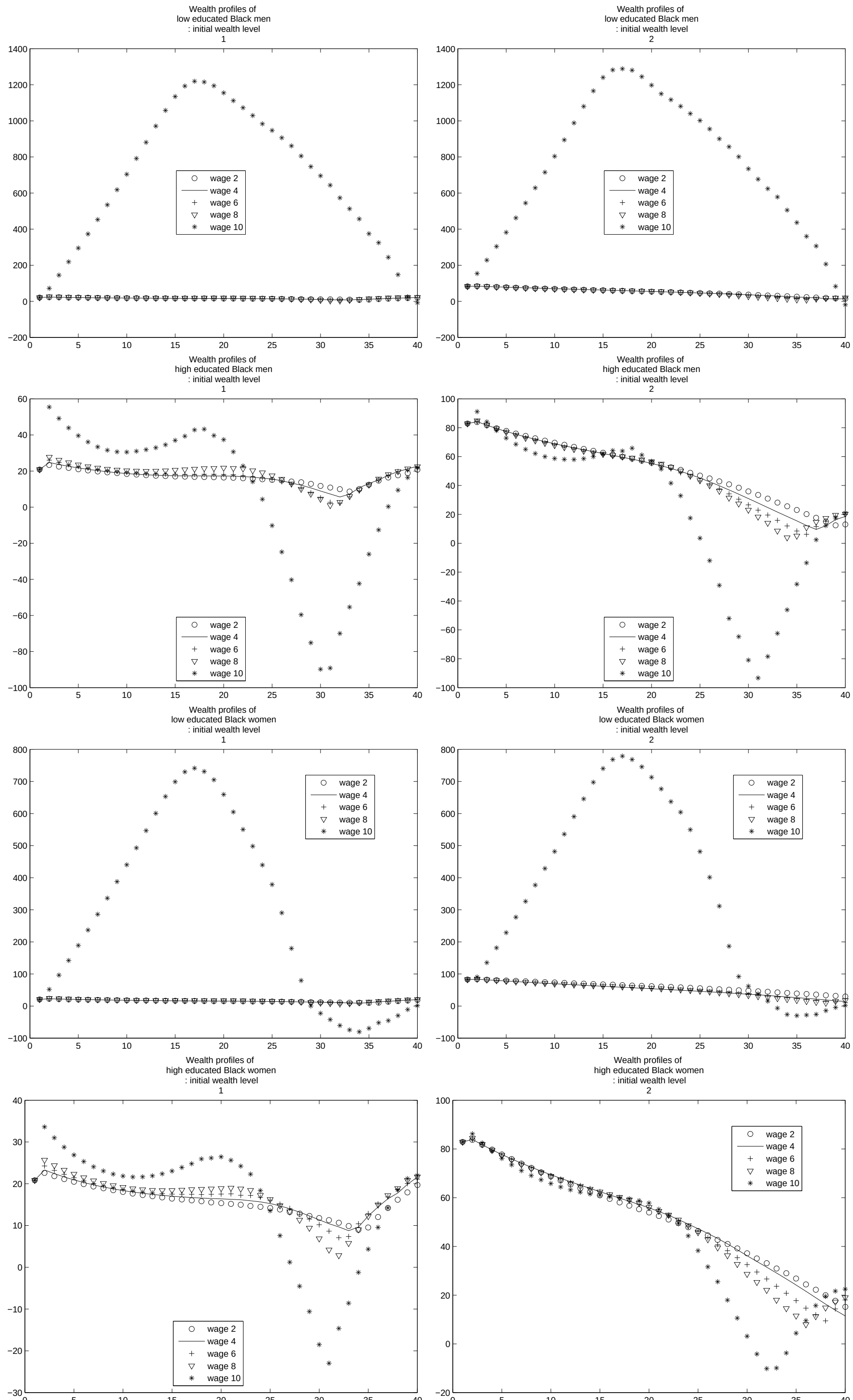


Figure 21: Wealth profiles in the model with uncertainty in wages

Appendix E

Chapter 2 Robustness checks

Table 9: OLS estimates of a happiness equation, conditioning on age, birth cohort and calendar time: first 3 months only

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Age	-0.0281***	0.0249**	0.1965***	0.2718***	0.1374***		-0.0461***	0.0009	0.1749***	0.1718***	0.023	
	(0.0060)	(0.0100)	(0.0300)	(0.0450)	(0.0350)		(0.0060)	(0.0090)	(0.0280)	(0.0380)	(0.0260)	
Age^2	0.0002***	-0.0003***	-0.0037***	-0.0041***	-0.0004***		0.0007***	0.0002**	-0.0039***	-0.0039***	0.0001	
	(0.0000)	(0.0000)	(0.0010)	(0.0010)	(0.0000)		(0.0000)	(0.0000)	(0.0010)	(0.0010)	(0.0000)	
Age^3			0.0000***	0.0000***					0.0000***	0.0000***		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Age 21						0.1481***						0.0947*
						(0.0570)						(0.0530)
Age 31						1.3983***						0.4551
						(0.3780)						(0.2880)
Age 41						2.4640***						0.7587
						(0.7200)						(0.5440)
Age 51						3.3642***						0.9651
						(1.0630)						(0.8030)
Age 60						4.4750***						1.4944
						(1.3720)						(1.0350)
[1929, 1934[		-0.1312						0.0093				
		(0.1070)						(0.0860)				
[1939, 1944[		-0.2393						-0.1593				
		(0.1500)						(0.1150)				
[1949, 1954[		-0.2151						-0.1607				
		(0.2230)						(0.1690)				

Continued on next page

Table 9 – continued from previous page

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
[1959, 1964[		-0.0711						-0.0555				
		(0.3010)						(0.2250)				
[1969, 1974[		0.2067						0.1769				
		(0.3770)						(0.2820)				
[1979, 1984[		0.2347						0.3182				
		(0.4530)						(0.3400)				
Cohort				0.0728**						-0.003		
				(0.0360)						(0.0280)		
$Cohort^2$			0.0011***	0.0008*					0.0002	0.0002		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
$Cohort^3$			0	0					0	0		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Born 1925					0.3859	0.462					0.0525	0.0675
					(0.2920)	(0.2930)					(0.2650)	(0.2650)
Born 1935					0.7129	0.9286**					-0.2773	-0.1695
					(0.4530)	(0.4530)					(0.3680)	(0.3680)
Born 1945					1.9611**	2.1981***					0.08	0.2043
					(0.7650)	(0.7640)					(0.5910)	(0.5910)
Born 1955					2.9769***	3.1545***					0.2088	0.2636
					(1.0950)	(1.0930)					(0.8360)	(0.8350)
Born 1965					4.2686***	4.4426***					0.587	0.632
					(1.4250)	(1.4240)					(1.0850)	(1.0840)
Born 1975					5.7526***	5.9642***					1.1567	1.236
					(1.7650)	(1.7630)					(1.3390)	(1.3370)
Born 1983					6.6240***	6.8774***					1.4314	1.5473
					(2.1570)	(2.1540)					(1.7390)	(1.7380)
Year 1986												

Continued on next page

Table 9 – continued from previous page

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Year 1991	-0.0746**	-0.1003*			-0.6225***	-0.6230***	0.0238	0.0075			-0.0743	-0.073
	(0.0370)	(0.0550)			(0.1760)	(0.1760)	(0.0340)	(0.0460)			(0.1350)	(0.1350)
Year 1996	-0.4659***	-0.5430***			-1.5998***	-1.6034***	-0.1104***	-0.1610**			-0.3373	-0.3384
	(0.0350)	(0.0890)			(0.3450)	(0.3440)	(0.0330)	(0.0690)			(0.2610)	(0.2610)
Year 2001	-0.3776***	-0.5185***			-2.1107***	-2.1149***	-0.1046***	-0.2090**			-0.4824	-0.4793
	(0.0360)	(0.1280)			(0.5160)	(0.5160)	(0.0340)	(0.0970)			(0.3900)	(0.3890)
Year			-0.1196***	-0.1857***					-0.0741***	-0.0714**		
			(0.0180)	(0.0380)					(0.0160)	(0.0300)		
$Year^2$			0.0017	0.002					0.0056***	0.0055***		
			(0.0020)	(0.0020)					(0.0010)	(0.0010)		
$Year^3$			0	0					-0.0002***	-0.0002***		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Constant	8.0463***	7.0126***	3.3584***	-0.5602	0.5025	2.8412**	4.9503***	4.0230***	1.2164**	1.3769	3.0493*	3.4010***
	(0.1170)	(0.4950)	(0.6150)	(2.0510)	(2.1050)	(1.4260)	(0.1470)	(0.3900)	(0.5180)	(1.5580)	(1.5960)	(1.0850)
R^2	0.017	0.019	0.016	0.016	0.021	0.022	0.297	0.299	0.298	0.298	0.3	0.301
Covariates	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Profile	Decreasing	Inverted U	Inverted U	Inverted U	Increasing	Increasing	U-shaped	Increasing	Inverted U	Inverted U	Unrelated	Unrelated
Inflexion Point		38.9	26.4	33.3			34.5		22.4	22.0		

Significance levels : $*$ 10% $**$ 5% $***$ 1% Standard errors in parentheses

Additional covariates are gender, *bundesland*, nationality, marital status, educational diploma, labor force status, household income and self reported satisfaction with health and number of members in the household.

Omitted categories: 20 year olds, year 1984, cohort born between $[1924, 1929[$, cohort born in 1924.

Table 10: OLS estimates of a happiness equation, conditioning on age, birth cohort and calendar time: first 4 months only

	I	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**
Age	-0.0290***	0.0201**	0.1649***	0.2396***	0.1239***		-0.0437***	0.0014	0.1674***	0.1606***	0.0115	
	(0.0060)	(0.0100)	(0.0280)	(0.0410)	(0.0310)		(0.0060)	(0.0080)	(0.0260)	(0.0350)	(0.0240)	
Age^2	0.0002***	-0.0003***	-0.0032***	-0.0036***	-0.0003***		0.0006***	0.0002**	-0.0039***	-0.0039***	0.0001*	
	(0.0000)	(0.0000)	(0.0010)	(0.0010)	(0.0000)		(0.0000)	(0.0000)	(0.0010)	(0.0010)	(0.0000)	
Age^3			0.0000***	0.0000***					0.0000***	0.0000***		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Age 21						0.1309**						0.068
						(0.0510)						(0.0480)
Age 31						1.2725***						0.3238
						(0.3410)						(0.2590)
Age 41						2.2678***						0.5211
						(0.6500)						(0.4900)
Age 51						3.0681***						0.5888
						(0.9600)						(0.7230)
Age 60						4.1266***						1.0325
						(1.2390)						(0.9320)
[1929, 1934[		-0.1005						0.0297				
		(0.1000)						(0.0810)				
[1939, 1944[		-0.1789						-0.1261				
		(0.1410)						(0.1100)				
[1949, 1954[		-0.1449						-0.1315				
		(0.2100)						(0.1610)				

Continued on next page

Table 10 – continued from previous page

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
[1959, 1964[		-0.0221						-0.0315				
		(0.2830)						(0.2150)				
[1969, 1974[		0.2439						0.1933				
		(0.3550)						(0.2680)				
[1979, 1984[		0.2655						0.3369				
		(0.4270)						(0.3240)				
Cohort				0.0725**						-0.0066		
				(0.0330)						(0.0250)		
Cohort^2			0.0009**	0.0005					-0.0001	0		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Cohort^3			0	0					0.0000**	0.0000*		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Born 1925					0.4254	0.5026*					-0.0339	-0.0129
					(0.2710)	(0.2720)					(0.2560)	(0.2570)
Born 1935					0.8065*	1.0247**					-0.3221	-0.2023
					(0.4130)	(0.4130)					(0.3420)	(0.3420)
Born 1945					1.9302***	2.1622***					-0.1205	0.0113
					(0.6940)	(0.6930)					(0.5400)	(0.5390)
Born 1955					2.8625***	3.0441***					-0.0956	-0.0281
					(0.9900)	(0.9890)					(0.7570)	(0.7560)
Born 1965					4.0929***	4.2810***					0.1785	0.2436
					(1.2890)	(1.2870)					(0.9810)	(0.9790)
Born 1975					5.4431***	5.6635***					0.6161	0.7128
					(1.5950)	(1.5940)					(1.2090)	(1.2070)
Born 1983					6.3422***	6.5932***					0.8088	0.9308
					(1.9300)	(1.9280)					(1.6110)	(1.6100)
Year 1986	0.1829***	0.1906***			0.3872***	0.3884***	0.1927***	0.1960***			0.2095***	0.2090***

Continued on next page

Table 10 – continued from previous page

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
	(0.0280)	(0.0320)			(0.0680)	(0.0680)	(0.0290)	(0.0310)			(0.0550)	(0.0550)
Year 1991	-0.2599***	-0.2799***			-0.5716***	-0.5749***	0.0328	0.0189			0.0017	-0.0005
	(0.0290)	(0.0380)			(0.0990)	(0.0990)	(0.0280)	(0.0340)			(0.0760)	(0.0760)
Year 1996	-0.2489***	-0.3208***			-1.1112***	-1.1178***	0.0867***	0.0405			-0.0164	-0.0213
	(0.0290)	(0.0690)			(0.2490)	(0.2490)	(0.0290)	(0.0550)			(0.1880)	(0.1880)
Year 2001	-0.1603***	-0.2933***			-1.5805***	-1.5889***	0.1033***	0.0052			-0.093	-0.095
	(0.0300)	(0.1060)			(0.4040)	(0.4040)	(0.0290)	(0.0810)			(0.3040)	(0.3040)
Year			-0.1642***	-0.2295***					-0.0966***	-0.0907***		
			(0.0160)	(0.0340)					(0.0150)	(0.0270)		
Year^2			0.0073***	0.0076***					0.0084***	0.0084***		
			(0.0010)	(0.0010)					(0.0010)	(0.0010)		
Year^3			-0.0002***	-0.0002***					-0.0003***	-0.0003***		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Constant	7.8630***	6.8628***	4.1916***	0.3116	0.5708	2.6710**	4.6810***	3.7818***	1.6625***	2.015	3.4716**	3.5931***
	(0.1080)	(0.4820)	(0.5770)	(1.8540)	(1.9640)	(1.3500)	(0.1390)	(0.3830)	(0.4910)	(1.4080)	(1.4860)	(1.0260)
R^2	0.016	0.017	0.015	0.015	0.019	0.02	0.292	0.293	0.293	0.293	0.294	0.295
Covariates	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Profile	Decreasing	Inverted U	Inverted U	Inverted U	Increasing	Increasing	U-shaped	Increasing	Inverted U	Inverted U	Increasing	Unrelated
Inflexion Point		36.6	25.8	33.6			34.1		21.3	20.7		

Significance levels : * 10% ** 5% *** 1% Standard errors in parentheses

Additional covariates are gender, *bundesland*, nationality, marital status, educational diploma, labor force status, household income and self reported satisfaction with health and number of members in the household.

Omitted categories: 20 year olds, year 1984, cohort born between $[1924, 1929[$, cohort born in 1924.

Table 11: OLS estimates of a happiness equation, conditioning on age, birth cohort and calendar time: first 6 months only

	I	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**
Age	-0.0294***	0.0205**	0.1594***	0.2465***	0.1331***		-0.0444***	0.0023	0.1632***	0.1578***	0.0149	
	(0.0060)	(0.0090)	(0.0270)	(0.0390)	(0.0290)		(0.0050)	(0.0080)	(0.0250)	(0.0330)	(0.0220)	
Age^2	0.0002***	-0.0003***	-0.0030***	-0.0035***	-0.0003***		0.0006***	0.0002**	-0.0037***	-0.0037***	0.0001*	
	(0.0000)	(0.0000)	(0.0010)	(0.0010)	(0.0000)		(0.0000)	(0.0000)	(0.0010)	(0.0010)	(0.0000)	
Age^3			0.0000***	0.0000***					0.0000***	0.0000***		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Age 21						0.1434***						0.0607
						(0.0480)						(0.0450)
Age 31						1.3746***						0.3529
						(0.3160)						(0.2390)
Age 41						2.4729***						0.5828
						(0.6030)						(0.4510)
Age 51						3.3837***						0.6813
						(0.8900)						(0.6660)
Age 60						4.5152***						1.158
						(1.1490)						(0.8580)
[1929, 1934[		-0.1256						0.0133				
		(0.0960)						(0.0790)				
[1939, 1944[		-0.182						-0.1379				
		(0.1360)						(0.1060)				
[1949, 1954[		-0.1312						-0.1276				
		(0.2030)						(0.1550)				

Continued on next page

Table 11 – continued from previous page

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
[1959, 1964[		-0.0016						-0.0206				
		(0.2740)						(0.2070)				
[1969, 1974[		0.2746						0.2177				
		(0.3430)						(0.2590)				
[1979, 1984[		0.3103						0.3719				
		(0.4130)						(0.3130)				
Cohort				0.0844***						-0.0052		
				(0.0300)						(0.0230)		
Cohort^2			0.0009***	0.0005					0	0.0001		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Cohort^3			0	0					0.0000**	0.0000*		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Born 1925					0.2931	0.347					-0.1104	-0.1003
					(0.2670)	(0.2680)					(0.2550)	(0.2560)
Born 1935					0.8781**	1.0844***					-0.2951	-0.1779
					(0.3900)	(0.3910)					(0.3250)	(0.3250)
Born 1945					2.0656***	2.2858***					-0.0821	0.0476
					(0.6480)	(0.6480)					(0.5030)	(0.5030)
Born 1955					3.0919***	3.2614***					-0.0255	0.0403
					(0.9210)	(0.9200)					(0.7010)	(0.7000)
Born 1965					4.4194***	4.5947***					0.2811	0.347
					(1.1970)	(1.1960)					(0.9060)	(0.9050)
Born 1975					5.8924***	6.0985***					0.7679	0.8629
					(1.4810)	(1.4800)					(1.1150)	(1.1140)
Born 1983					6.8449***	7.0807***					0.8868	1.007
					(1.7450)	(1.7430)					(1.3750)	(1.3730)
Year 1986	0.1637***	0.1783***			0.5048***	0.5073***	0.1314***	0.1388***			0.1686**	0.1683**

Continued on next page

Table 11 – continued from previous page

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
	(0.0270)	(0.0350)			(0.0910)	(0.0910)	(0.0280)	(0.0320)			(0.0700)	(0.0700)
Year 1991	-0.2413***	-0.2603***			-0.4702***	-0.4714***	-0.0211	-0.0356			-0.0502	-0.0518
	(0.0270)	(0.0310)			(0.0640)	(0.0640)	(0.0260)	(0.0290)			(0.0510)	(0.0510)
Year 1996	-0.2512***	-0.3286***			-1.0817***	-1.0865***	0.0139	-0.0386			-0.1056	-0.111
	(0.0280)	(0.0590)			(0.2030)	(0.2020)	(0.0270)	(0.0480)			(0.1530)	(0.1530)
Year 2001	-0.1665***	-0.3116***			-1.6060***	-1.6123***	0.0368	-0.074			-0.1946	-0.1975
	(0.0290)	(0.0950)			(0.3460)	(0.3460)	(0.0280)	(0.0730)			(0.2590)	(0.2590)
Year			-0.1503***	-0.2252***					-0.0873***	-0.0827***		
			(0.0160)	(0.0320)					(0.0140)	(0.0250)		
$Year^2$			0.0061***	0.0063***					0.0071***	0.0071***		
			(0.0010)	(0.0010)					(0.0010)	(0.0010)		
$Year^3$			-0.0001***	-0.0001***					-0.0002***	-0.0002***		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Constant	7.8773***	6.8290***	4.1277***	-0.3556	-0.0777	2.2236*	4.7480***	3.7981***	1.5692***	1.8447	3.3249**	3.5256***
	(0.1040)	(0.4730)	(0.5580)	(1.7200)	(1.8510)	(1.2830)	(0.1330)	(0.3740)	(0.4730)	(1.2990)	(1.3920)	(0.9690)
R^2	0.015	0.016	0.014	0.015	0.018	0.019	0.291	0.293	0.292	0.292	0.293	0.294
Covariates	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Profile	Decreasing	Inverted U	Inverted U	Inverted U	Increasing	Increasing	U-shaped	Increasing	Inverted U	Inverted U	Increasing	Unrelated
Inflexion Point	66.9	38.5	26.6	35.7			34.3		21.8	21.3		

Significance levels : * 10% ** 5% *** 1% Standard errors in parentheses

Additional covariates are gender, *bundesland*, nationality, marital status, educational diploma, labor force status, household income and self reported satisfaction with health and number of members in the household.

Omitted categories: 20 year olds, year 1984, cohort born between [1924, 1929[, cohort born in 1924.

Table 12: OLS estimates of a happiness equation, conditioning on age, birth cohort and calendar time: balanced panel

	I	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**
Age	-0.0297***	0.0219	0.1666***	0.1289***	0.0193		-0.0455***	0.008	0.1289***	0.0865**	-0.0069	
	(0.0080)	(0.0140)	(0.0380)	(0.0500)	(0.0360)		(0.0080)	(0.0120)	(0.0350)	(0.0440)	(0.0280)	
Age^2	0.0003***	-0.0002**	-0.0033***	-0.0030***	-0.0003***		0.0007***	0.0002*	-0.0025***	-0.0022**	0.0001	
	(0.0000)	(0.0000)	(0.0010)	(0.0010)	(0.0000)		(0.0000)	(0.0000)	(0.0010)	(0.0010)	(0.0000)	
Age^3			0.0000***	0.0000***					0.0000***	0.0000***		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Age 21						0.0701						0.0929
						(0.0690)						(0.0640)
Age 31						0.125						0.1084
						(0.3910)						(0.2950)
Age 41						0.1856						0.1419
						(0.7430)						(0.5530)
Age 51						0.0463						0.0975
						(1.0980)						(0.8140)
Age 60						0.0991						0.2526
						(1.4150)						(1.0470)
$[1929, 1934[$		-0.1573						-0.0801				
		(0.1390)						(0.1070)				
$[1939, 1944[$		-0.2686						-0.1461				
		(0.2060)						(0.1530)				
$[1949, 1954[$		-0.1773						-0.0762				
		(0.3120)						(0.2310)				

Continued on next page

Table 12 – continued from previous page

	I	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**
[1959, 1964[		0.0605						0.1722				
		(0.4290)						(0.3140)				
[1969, 1974[		0.2964						0.4499				
		(0.5410)						(0.3970)				
[1979, 1984[		0.403						0.6564				
		(0.6520)						(0.4800)				
Cohort				-0.0363						-0.0409		
				(0.0380)						(0.0290)		
Cohort^2			0.0007	0.0010*					0.0004	0.0007		
			(0.0010)	(0.0010)					(0.0000)	(0.0000)		
Cohort^3			0	0					0	0		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Born 1925					-0.2031	-0.1726					-0.1295	-0.1303
					(0.4290)	(0.4300)					(0.3490)	(0.3500)
Born 1935					-0.7675	-0.6490					-0.6964	-0.6684
					(0.5370)	(0.5380)					(0.4250)	(0.4260)
Born 1945					-0.7802	-0.6533)					-0.7664	-0.7331
					(0.8290)	(0.8290)					(0.6270)	(0.6270)
Born 1955					-0.6372	-0.5482					-0.6883	-0.6867
					(1.1550)	(1.1550)					(0.8630)	(0.8630)
Born 1965					-0.6702	-0.5727					-0.7785	-0.7738
					(1.4860)	(1.4860)					(1.1100)	(1.1100)
Born 1975					-0.167	-0.0489					-0.4483	-0.4248
					(1.8340)	(1.8340)					(1.3620)	(1.3620)
Born 1983					0.5771	0.7229					0.3735	0.4284
					(2.1140)	(2.1120)					(1.6300)	(1.6290)
Year 1986	0.4921***	0.014			0.5902	0.5922	0.3055***	0.5297***			0.2347	0.2343

Continued on next page

Table 12 – continued from previous page

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
	(0.0420)	(0.0370)			(0.5990)	(0.5990)	(0.0390)	(0.1430)			(0.4170)	(0.4170)
Year 1991	0.5215***	0.0023			0.6244	0.6241	0.4016***	0.5722***			0.3819	0.3806
	(0.0360)	(0.0810)			(0.4210)	(0.4220)	(0.0340)	(0.1000)			(0.2860)	(0.2860)
Year 1996	0.1745***	-0.4055***			0.234	0.2313	0.1093***	0.2104***			0.1012	0.0969
	(0.0350)	(0.1370)			(0.2480)	(0.2480)	(0.0330)	(0.0610)			(0.1590)	(0.1590)
Year 2001	0.2316***	-0.4177**			0.2602***	0.2589***	0.1169***	0.1359***			0.1222***	0.1211***
	(0.0310)	(0.1950)			(0.0770)	(0.0770)	(0.0280)	(0.0290)			(0.0380)	(0.0380)
Year			-0.0172	0.0127					0	0.0337		
			(0.0210)	(0.0400)					(0.0200)	(0.0310)		
Year^2			-0.0036**	-0.0038**					-0.0026	-0.0027*		
			(0.0020)	(0.0020)					(0.0020)	(0.0020)		
Year^3			0.0001*	0.0001*					0	0		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Constant	7.5589***	6.9586***	3.9571***	5.7902***	7.1384**	7.2771***	4.5953***	3.1212***	1.3530*	3.4180**	4.7178**	4.5645***
	(0.1580)	(0.7020)	(0.8480)	(2.1250)	(2.7810)	(2.0790)	(0.2150)	(0.6880)	(0.7000)	(1.6280)	(2.0520)	(1.5270)
R^2	0.011	0.013	0.01	0.01	0.017	0.018	0.297	0.299	0.296	0.296	0.301	0.302
Covariates	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Profile	U-shaped	Decreasing	Inverted U	Inverted U	Decreasing	Unrelated	U-shaped	Increasing	Inverted U	Inverted U	Unrelated	Unrelated
Inflexion Point	53.7		25.4	21.4			33.7		25.4	19.3		

Significance levels : * 10% ** 5% *** 1% Standard errors in parentheses

Additional covariates are gender, *bundesland*, nationality, marital status, educational diploma, labor force status, household income and self reported satisfaction with health and number of members in the household.

Omitted categories: 20 year olds, year 1984, cohort born between $[1924, 1929[$, cohort born in 1924.

Table 13: OLS estimates of a happiness equation, conditioning on age, birth cohort and calendar time: individuals who answer in first and last waves

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Age	-0.0297***	0.0223	0.1667***	0.1261**	0.0162		-0.0451***	0.0084	0.1292***	0.0854*	-0.0081	
	(0.0080)	(0.0140)	(0.0380)	(0.0500)	(0.0360)		(0.0080)	(0.0120)	(0.0350)	(0.0440)	(0.0280)	
Age^2	0.0003***	-0.0002**	-0.0033***	-0.0030***	-0.0003***		0.0007***	0.0002*	-0.0026***	-0.0022**	0.0001	
	(0.0000)	(0.0000)	(0.0010)	(0.0010)	(0.0000)		(0.0000)	(0.0000)	(0.0010)	(0.0010)	(0.0000)	
Age^3			0.0000***	0.0000***					0.0000***	0.0000***		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Age 21						0.0705						0.0936
						(0.0690)						(0.0640)
Age 31						0.095						0.0939
						(0.3910)						(0.2950)
Age 41						0.1302						0.1161
						(0.7420)						(0.5520)
Age 51						-0.0419						0.0531
						(1.0970)						(0.8120)
Age 60						-0.0113						0.1966
						(1.4130)						(1.0450)
[1929, 1934[		-0.1579						-0.0809				
		(0.1390)						(0.1070)				
[1939, 1944[		-0.2661						-0.1477				
		(0.2060)						(0.1530)				
[1949, 1954[		-0.172						-0.0788				
		(0.3120)						(0.2310)				

Continued on next page

Table 13 – continued from previous page

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
[1959, 1964[		0.0668						0.1644				
		(0.4280)						(0.3140)				
[1969, 1974[		0.3042						0.444				
		(0.5410)						(0.3960)				
[1979, 1984[		0.4204						0.6537				
		(0.6520)						(0.4790)				
Cohort				-0.0391						-0.0422		
				(0.0380)						(0.0290)		
Cohort^2			0.0007	0.0010*					0.0004	0.0007		
			(0.0010)	(0.0010)					(0.0000)	(0.0000)		
Cohort^3			0	0					0	0		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Born 1925					-0.2339	(0.2035)					-0.1289	(0.1296)
					(0.4220)	(0.4230)					(0.3490)	(0.3500)
Born 1935					-0.8245	(0.7063)					-0.7117*	(0.6828)
					(0.5310)	(0.5320)					(0.4250)	(0.4250)
Born 1945					-0.865	(0.7395)					-0.7936	(0.7603)
					(0.8250)	(0.8250)					(0.6260)	(0.6260)
Born 1955					-0.7499	-0.6629					-0.7331	-0.7323
					(1.1510)	(1.1510)					(0.8620)	(0.8610)
Born 1965					-0.814	-0.719					-0.8367	-0.833
					(1.4820)	(1.4830)					(1.1080)	(1.1080)
Born 1975					-0.347	-0.2328					-0.5257	-0.5039
					(1.8290)	(1.8290)					(1.3580)	(1.3580)
Born 1983					0.3882	0.5342					0.2892	0.3455
					(2.1100)	(2.1090)					(1.6270)	(1.6260)
Year 1986	0.4052***	0.0112			0.4447	0.4459	0.3061***	0.5287***			0.2118	0.2105

Continued on next page

Table 13 – continued from previous page

	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
	(0.0410)	(0.0370)			(0.5630)	(0.5630)	(0.0390)	(0.1430)			(0.4160)	(0.4160)
Year 1991	0.4356***	-0.0007			0.4948	0.494	0.4028***	0.5725***			0.3672	0.3654
	(0.0350)	(0.0810)			(0.3860)	(0.3860)	(0.0340)	(0.1000)			(0.2850)	(0.2860)
Year 1996	0.0896***	-0.4092***			0.1184	0.1155	0.1119***	0.2125***			0.0948	0.0902
	(0.0330)	(0.1360)			(0.2130)	(0.2130)	(0.0330)	(0.0610)			(0.1580)	(0.1580)
Year 2001	0.1463***	-0.4246**			0.1573***	0.1562***	0.1177***	0.1363***			0.1208***	0.1197***
	(0.0270)	(0.1940)			(0.0450)	(0.0450)	(0.0280)	(0.0290)			(0.0380)	(0.0380)
Year			-0.0185	0.0136					-0.0001	0.0347		
			(0.0210)	(0.0390)					(0.0200)	(0.0310)		
Year^2			-0.0034**	-0.0036**					-0.0025	-0.0027*		
			(0.0020)	(0.0020)					(0.0020)	(0.0020)		
Year^3			0.0001*	0.0001*					0	0		
			(0.0000)	(0.0000)					(0.0000)	(0.0000)		
Constant	7.6446***	6.9419***	3.9836***	5.9555***	7.4857***	7.5638***	4.5893***	3.1218***	1.3659*	3.4966**	4.8255**	4.6461***
	(0.1570)	(0.7020)	(0.8460)	(2.1230)	(2.7420)	(2.0410)	(0.2140)	(0.6870)	(0.6990)	(1.6240)	(2.0470)	(1.5240)
R^2	0.011	0.013	0.01	0.01	0.017	0.018	0.296	0.298	0.296	0.296	0.301	0.302
Covariates	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Profile	U-shaped	Decreasing	Inverted U	Inverted U	Decreasing	Unrelated	U-shaped	Increasing	Inverted U	Inverted U	Unrelated	Unrelated
Inflexion Point	53.7		25.3	21.0			33.6		25.3	19.1		

Significance levels : $*$ 10% $**$ 5% $***$ 1% Standard errors in parentheses

Additional covariates are gender, *bundesland*, nationality, marital status, educational diploma, labor force status, household income and self reported satisfaction with health and number of members in the household.

Omitted categories: 20 year olds, year 1984, cohort born between $[1924, 1929[$, cohort born in 1924.

Table 14: Accounting for unobserved heterogeneity and the ordinal nature of the dependent variable

	Within Groups		**Probit**		**Ordered Fixed Effects Logit**	
Age	-0.0714***		0.0026		-0.1180***	
	(0.0210)		(0.0140)		(0.0320)	
Age^2	0.0002**		0.0001		0.0003***	
	(0.0000)		(0.0000)		(0.0000)	
Age 21		-0.0244		0.028		-0.0314
		(0.0440)		(0.0300)		(0.0700)
Age 31		-0.5821***		0.1302		-1.0125***
		(0.2210)		(0.1550)		(0.3450)
Age 41		-1.1590***		0.1882		-1.9226***
		(0.4150)		(0.2930)		(0.6500)
Age 51		-1.9002***		0.184		-3.0668***
		(0.6120)		(0.4330)		(0.9570)
Age 60		-2.1504***		0.3957		-3.5207***
		(0.7880)		(0.5580)		(1.2340)
Born 1925			-0.1096	-0.097		
			(0.1840)	(0.1850)		
Born 1935			-0.3651*	-0.2927		
			(0.2200)	(0.2200)		
Born 1945			-0.2924	-0.2125		
			(0.3320)	(0.3320)		
Born 1955			-0.3559	-0.3124		
			(0.4590)	(0.4590)		
Born 1965			-0.2391	-0.1928		
			(0.5900)	(0.5900)		
Born 1975			0.0024	0.0648		
			(0.7260)	(0.7260)		
Born 1983			0.1888	0.258		
			(0.9050)	(0.9040)		
Year 1986	-0.6473*	-0.6355*	0.3266	0.3278	-1.0050*	-0.9789*
	(0.3330)	(0.3340)	(0.2380)	(0.2380)	(0.5220)	(0.5230)
Year 1991	-0.4452*	-0.4354*	0.202	0.2022	-0.6506*	-0.6298*
	(0.2350)	(0.2350)	(0.1670)	(0.1670)	(0.3680)	(0.3680)
Year 1996	-0.1684	-0.1665	0.1710*	0.1688*	-0.28	-0.2723
	(0.1390)	(0.1390)	(0.0990)	(0.0990)	(0.2180)	(0.2180)
Year 2001	0.1065**	0.1054**	0.1506***	0.1489***	0.2019***	0.2011***
	(0.0450)	(0.0450)	(0.0320)	(0.0320)	(0.0730)	(0.0730)
Constant	7.6234***	6.1821***				

Continued on next page

Table 14 – continued from previous page

	Within Groups		**Probit**		**Ordered Fixed Effects Logit**	
	(0.9680)	(0.5840)				
R^2	0.114	0.116				
Covariates	Yes	Yes	Yes	Yes	Yes	Yes
Profile	Decreasing	Decreasing	Unrelated	Unrelated	Decreasing	Decreasing

Significance levels : $*$: 10% $**$: 5% $***$: 1%

Robust standard errors in parentheses.

Additional covariates are gender, *bundesland*, nationality, marital status, educational diploma, labor force status, household income and self-reported satisfaction with health and number of members in the household.

Omitted categories: 20 year olds, year 1984, cohort born between [1924, 1929[, cohort born in 1924.

Table 15: Using schooling cohorts and higher frequency intervals for age and period

	yearly frequency	**Monthly Frequency**
Age 21	-0.0233	
	(0.0428)	
Age 31	-0.6052***	
	(0.1963)	
Age 41	-1.2799***	
	(0.3679)	
Age 51	-2.0491***	
	(0.5414)	
Age 60	-2.3884***	
	(0.6979)	
Cohort 1931	1.9868	0.9454***
	(1.2e+03)	(0.3168)
Cohort 1941	1.0645	0.4980**
	(2.1e+03)	(0.2361)
Cohort 1951	0.1720	0.2971
	(1.2e+03)	(0.2272)
Cohort 1961	-0.6329	0.0893
	(789.3656)	(0.2241)
Cohort 1971	-1.2196	0.1074
	(2.3e+03)	(0.2192)

Continued on next page

Table 15 – continued from previous page

	yearly frequency	Monthly Frequency
Cohort 1981	-1.6821	0.3503
	(2.1e+03)	(0.2195)
Cohort 1989	-2.2525	
	(1.2e+03)	
Year 1986	-0.3309***	
	(0.0738)	
Year 1991	-0.1050***	
	(0.0296)	
Year 1996	0.2649**	
	(0.1065)	
Year 2001	0.6195***	
	(0.1913)	
R^2	0.2942	0.2989
N	132387	93253

Significance levels : $*$: 10% $**$: 5% $***$: 1%

Robust standard errors in parentheses.

Additional covariates are gender, *bundesland*, nationality, marital status, educational diploma, labor force status, household income and self-reported satisfaction with health and number of members in the household.

Omitted categories: 20 year olds, year 1984, cohort in school in 1929.

Appendix F

Chapter 3 data

The variables used in the paper are as follows.

The Measure of Life Satisfaction (Happiness)

[General Satisfaction] How satisfied or dissatisfied are you with your life as a whole? (1 completely satisfied up to 7 completely dissatisfied).

The Measures of Experienced Domestic Violence

The data set contains two main variables on experienced domestic violence.

Have you ever been a victim of domestic violence (yes=1/no=0) (Domestic Violence ever)

and

Please indicate how vulnerable you feel to domestic violence in the future - using a scale of 1 to 7 where 1 means "not at all vulnerable" and 7 means "very vulnerable"?

The latter variable was turned into a binary variable, taking the value 1 when vulnerability was 4 or higher.

Socio-economic and demographic variables

[Age]

[Gender] (1 male 2 female)

[Household Income$_p$] A continuous measure based on BHPS annual household income variable, calculated by replacing each income band value with the income of the BHPS variable, averaged over the values within the income band.

[Household Income$_i$] Imputed household income based on BHPS 2004 data, by matching observations according to Age and Age squared, Gender, Ethnicity, Marital Status, Education Attained, Employment Status, Dependent Children, Religion, Regional dummies, Individual income brackets. Interval data estimation was used. The model imposes by construction that the imputed value falls within the respondent's household income bracket; however, the probability that that is so is very low for most of the observations. These probabilities vary from 0.013 for the first percentile, is around one third for the third quartile, two thirds for the ninth decile and only reaches 0.943 at the 95% percentile.

[Ethnicity] (1 White British 2 Non-white British)

[Marital Status] What is your marital status? (1 married or living as married 2 separated or divorced 3 widowed or never married)

[Education Attained] What is the highest educational or work-related qualification you have? (too many options and regional differences these were later collapsed into 4 comparable categories 1 Other Schooling 2 Vocational Diploma 3 CSE A Level 4 University Degree)

[Employment Status] Which of these best applies to you? (1 working 8 or more hours per week 2 working less than 8 hours per week)

[Dependent Children] How many dependent children do you have that is children dependent on your income? (1 "none" 2 "at least one")

[Log Hourly wage] Potential wage estimated using comparable individuals from BHPS. Individuals were matched on the following observables: Age and Age squared, Gender, Year dummies (from 1998 to 2004), Ethnicity, Marital Status, Education Attained, Employment Status, Dependent Children, Religion and Regional dummies.

[Postcode and Local Crime Rates] Can you tell us the first part of your postcode this can include up to four letters and numbers (e.g. SE23)? Crime rates were then retrieved based on postcode information. Local crime data were collected online from http://www.crimestatistics.org.uk/tool/. This variable measures the number of all reported crime offences per 1000 individuals in the first quarter of 2004. It is collected at the CDRP (Crime and Disorder Reduction Partnerships) level, throughout England and Wales only (we hence lost the 90 observations corresponding to the Scottish sample). It combines police records with the British Crime Survey self-reported questionnaire of individual experiences.

The Measure of Personality

The measure of personality used derives from answers to the ten questions below. Each personality dimension combines two polarised traits, so that the positive one enters positively and the negative one enters negatively towards the final score. The score for each of the five dimensions is then based on the difference between the two relevant traits (the former minus the latter) and can take a value in the range from -6 to 6.

[Extraversion] (-6 up 6)

(+) I see myself as extraverted, enthusiastic (1 agree strongly up to 7 disagree strongly)

(–) I see myself as reserved, quiet (1 agree strongly up to 7 disagree strongly)

[Agreeableness] (-6 up 6)

(+) I see myself as sympathetic, warm (1 agree strongly up to 7 disagree strongly)

(–) I see myself as critical, quarrelsome (1 agree strongly up to 7 disagree strongly)

[Conscientiousness] (-6 up 6)

(+) I see myself as dependable, self-disciplined (1 agree strongly up to 7 disagree strongly)

(–) I see myself as disorganised, careless (1 agree strongly up to 7 disagree strongly)

[Emotional Stability] (-6 up 6)

(+) I see myself as calm, emotionally stable (1 agree strongly up to 7 disagree strongly)

(–) I see myself as anxious, easily upset (1 agree strongly up to 7 disagree strongly)

[Openness] (-6 up 6)

(+) I see myself as open to new experience, complex (1 agree strongly up to 7 disagree strongly)

(–) I see myself as conventional, uncreative (1 agree strongly up to 7 disagree strongly)

The Measure of Distress

The distress index was based on the following questions.

Have you recently lost much sleep over worry? 1. Not at all 2. No more than usual 3. Rather more than usual 4. Much more than usual

Have you recently felt constantly under strain? 1. Not at all 2. No more than usual 3. Rather more than usual 4. Much more than usual

Have you recently been able to enjoy your normal day-to- day activities? 1. More so than usual 2. Same as usual 3. Less so than usual 4. Much less than usual

At work, have you recently felt that you were playing a useful part in things? 1. More so than usual 2. Same as usual 3. Less so than usual 4. Much less than usual

Outside of work, have you recently felt that you were playing a useful part in things? 1. More so than usual 2. Same as usual 3. Less so than usual 4. Much less than usual

Have you recently been enjoying your recreational activities? 1. More so than usual 2. Same as usual 3. Less so than usual 4. Much less than usual

Have you recently been thinking of yourself as a worthless person? 1. Not at all 2. No more than usual 3. Rather more than usual 4. Much more than usual

All variables were turned into binary variables, where the value 1 indicates distress (previous values 3 and 4). All of these values were then added up to create the index.

Bibliography

Agarwal, B. (2006). Towards freedom from domestic violence: The neglected obvious. In Groeningen (Ed.), *International Conference of the HDCA: Freedom and Justice*.

Aizer, A. (2007). Wages, violence and health in the household. Working Paper 13494, NBER.

Anand, P., Hunter, G., Carter, I., Dowding, K., Guala, F., & Hees, M. V. (2009). The development of capability indicators. *Journal of Human Development and Capabilities*, *10*(1), 125.

Anand, P. & Santos, C. (2007). Violent crime, gender inequalities and well-being: models based on a survey of individual capabilities and crime rates for england and wales. *Revue d'Economie Politique*, *117*(1), 135.

Atkinson, G., Healey, A., & Mourato, S. (2005). Valuing the costs of violent crime: a stated preference approach. *Oxford Economic Papers*, *57*, 559.

B. Fitzenberger, R. Hujer, T. M. & Schnabel, R. (2001). Testing for uniform wage trends in west germany: a cohort analysis using quantile regressions for censored data. *Empirical Economics*, *26*, 41.

Becketti, S., Gould, W., Lillard, L., & Welch, F. (1988). The panel study of income dynamics after fourteen years: An evaluation. *Journal of Labor Economics*, *6*(4), 472–492.

Berger, M. C. (1985). The effect of cohort size on earnings growth: a reexamination of the evidence. *The Journal of Political Economy*, *93*(3), 561.

Blundell, R., Gosling, A., Ichimura, H., & Meghir, C. (2007). Changes in the distribution of male and female wages accounting for employment composition using bounds. *Econometrica*, *75*(2), 323–363.

Blundell, R., Reed, H., & Stoker, T. M. (2003). Interpreting aggregate wage growth: The role of labor market participation. *American Economic Review*, *93*(4).

Bowlus, A. J. & Seitz, S. (2006). Domestic violence, employment and divorce. *International Economic Review*, *47*(4), 1113.

Brand, S. & Price, R. (2001). The economic and social costs of crime. Research Study 217, Home Office, London, UK.

Card, D. & Lemieux, T. (2000). Dropout and enrolment trends in the post-war period: what went wrong in the 1970's? Working Paper 7658, NBER.

Card, D. & Lemieux, T. (2001). Can falling supply explain the rising return to college for younger men? a cohort-based analysis. *Quarterly Journal of Economics*, *May*, 705.

Carroll, N., Frijters, P., & Shields, M. A. Quantifying the costs of drought: new evidence from life satisfaction data. *Journal of Population Economics*, (forthcoming).

Chandra, A. (2003). Is the convergence in the racial wage gap illusory? Working Paper 9476, NBER.

Chaplin, R., Flatley, J., & Smith, K. (2011). Crime in england and wales 2010/11 findings from the british crime survey and police recorded crime, second edition. Home Office Statistical Bulletin 10, Home Office.

Clark, A., Frijters, P., & Shields, M. A. (2006). Income and happiness: Evidence, explanations and economic implications. Working Paper 24, Paris-Jourdan Sciences Economiques.

Clark, A. E. (2002). Born to be mild: cohort effects in subjective well-being. *mimeo*.

Clements, M. S., Armstrong, B. K., & Moolgavkar, S. H. (2005). Lung cancer rate predictions using generalized additive models. *Biostatistics*, *6*(4), 576.

Cribier, F. (2005). Changes in the experiences of life between two cohorts of parisian pensioners, born in circa 1907 and 1921. *Ageing and Society*, *25*, 637.

Cummins, R. A. (1996). The domains of life satisfaction: an attempt to order chaos. *Social Indicators Research*, *38*, 303.

Deaton, A., Fortson, J., & Tortora, R. (2009). Life (evaluation), hiv/aids, and death in africa. Working Paper 14637, NBER.

Deaton, A. & Paxson, C. (1999). Growth, demographic structure, and national saving in taiwan. *mimeo*.

Dolan, P. & Metcalfe, R. (2008). Valuing non-market goods: A comparison of preference-based and experience-based approaches. mimeo.

Dye, J. & Sosimi, J. (2006). United kingdom national accounts, the blue book. 2006 edition, Office for National Statistics.

Easterlin, R. (1974). Does economic growth improve the human lot? some empirical evidence. In R. David & R. Reder (Eds.), *Nations and Households in Economic Growth: Essays in Honor of Moses Abramovitz*. New York: Academic Press.

Easterlin, R. (2001). Income and happiness: towards a unified theory. *Economic Journal*, *111*, 465.

Eckstein, Z. & Wolpin, K. I. (1989). Dynamic labour force participation of married women and endogenous work experience. *The Review of Economic Studies*, *56*(3), 375–390.

Farmer, A. & Tiefenthaler, J. (1997). An economic analysis of domestic violence. *Revue of Social Economy*, *55*(3), 337.

Ferrer-i-Carbonell, A. & Frijters, P. (2004). How important is methodology for the estimates of the determinants of happiness? *The Economic Journal*, *114*, 641.

Fitzgerald, J., Gottschalk, P., & Moffitt, R. (1998). An analysis of sample attrition in panel data: The michigan panel study of income dynamics. *The Journal of Human Resources*, *33*(2), 251–299.

Fortin, N. M. (2006). Greed, altruism and the gender wage gap. *mimeo*.

Frey, B. S., Luechinger, S., & Stutzer, A. (2004). Valuing public goods: The life satisfaction approach. Working Paper 11, CREMA, Basel, Switzerland.

Frey, B. S. & Stutzer, A. (2002). What can economists learn from happiness research? *Journal of Economic Literature*, *40*(2), 402.

Frijters, P. & Beatton, T. (2008). The mystery of the u-shaped relationship between happiness and age. Working Paper 26, NCER Working Paper Series.

Gibbons, S. & Machin, S. (2008). Valuing school quality, better transport, and lower crime: evidence from house prices. *Oxford Review of Economic Policy*, *24*(1), 99.

Gosling, S. D., Rentfrow, P. J., & Jr., W. B. S. (2003). A very brief measure of the big-five personality domains. *Journal of Research in Personality*, *37*, 504.

Gottschalk, P. & Moffitt, R. A. (1992). Earnings and wage distributions in the nls, cps, and psid. report to the u.s. department of labor, Brown University.

Hall, B. H., Mairesse, J., & Turner, L. (2005). Identifying age, cohort and period effects in scientific research productivity: discussion and illustration using simulated and actual data on french physicists. *mimeo*.

Hall, P. & Smith, K. (2011). Analysis of the 2010/11 british crime survey intimate personal violence splitsample experiment. Home office methodology report, Home Office.

Handcock, M. S., Morris, M., & Bernhardt, A. (2000). Comparing earnings inequality using two major surveys. *Monthly Labor Review*, (March), 48–61.

Heckman, J. (1974a). Life cycle consumption and labor supply: An explanation of the relationship between income and consumption over the life cycle. *American Economic Review*, *64*(1), 188–194.

Heckman, J. (1974b). Shadow prices, market wages and labour supply. *Econometrica*, *42*(4), 679–694.

Heckman, J., Lyons, T., & Todd, P. (2000). Understanding black-white wage differentials, 1960-1990. *American Economic Review*, *90*(2), 344–349.

Heckman, J. & Robb, R. (1985). Using longitudinal data to estimate age, period and cohort effects in earnings equations. In S. Feinberg & W. Mason (Eds.), *Analyzing Longitudinal Data to Estimate Age, Period and Cohort Effects*. New York: Academic Press.

Hidrobo, M. & Fernald, L. (2013). Cash transfers and domestic violence. *Journal of Health Economics*, *32*(1), 304.

Holford, T. R. (1983). The estimation of age, period and cohort effects for vital rates. *Biometrics*, *39*, 311.

Jarvinen, J., Kail, A., & Miller, I. (2008). Hard knock life: Violence against women, a guide for donors and funders. Report, New Philanthropy Capital, London, UK.

Juhn, C., Murphy, K., & Pierce, B. (1993). Wage inequality and the rise in returns to skill. *Journal of Political Economy*, *101*(3), 410–442.

Jurges, H. (2003). Age, cohort, and the slump in job satisfaction among west german workers. *Labour*, *17*(4), 489.

Kahneman, D. & Krueger, A. B. (2006). Developments in the measurement of subjective well-being. *Journal of Economic Perspectives*, *20*(1), 3.

Kermack, W. O., McKendrick, A. G., & McKinley, P. L. (1934). Death rates in great britain and sweden: expression of specific mortality rates as products of 2 factors, and some consequences thereof. *Journal of Hygiene*, *3334*, 433.

Knetsch, J. (2000). Environmental valuations and standard theory: behavioral findings, context dependence and implications. In T. Tietenberg & H. Folmer (Eds.), *The International Yearbook of Environmental and Resource Economics*. Cheltenham, UK: Edward Elgar.

Layard, R., Mayraz, G., & Nickell, S. (2007). The marginal utility of income. Discussion Paper 784, CEP.

Low, H. W. (2005). Self-insurance in a life-cycle model of labour supply and savings. *Review of Economic Dynamics*, *8*, 945–975.

Lucas, R. E., Clark, A. E., Georgellis, Y., & Diener, E. (2003). Reexamining adaptation and the set point model of happiness: Reactions to changes in marital status. *Journal of Personality and Social Psychology*, *84*(3), 527.

Luechinger, S. (2009). Valuing air quality using the life satisfaction approach. *The Economic Journal*, *119*, 482.

Lundberg, S. (2010). Personality and marital surplus. *mimeo*.

Luttmer, E. F. P. (2005). Neighbors as negatives: relative earnings and well-being. *Quarterly Journal of Economics*, 963.

MaCurdy, T. & Mroz, T. (1995). Measuring macroeconomic shifts in wages from cohort specifications. *mimeo*.

Moreno, C. G., Jansen, H., Ellsberg, M., Heise, L., & Watts, C. (2005). Multi-country study on womens health and domestic violence against women : initial results on prevalence, health outcomes and womens responses. report, World Health Organization, Geneva, Switzerland.

Morrison, A. & Biehl, M. L. (1999). *Too Close to Home: Domestic Violence in the Americas*. Washington, DC: IDB.

Mulligan, C. B. & Rubinstein, Y. (2004). The closing of the gender gap as a roy model illusion. Working Paper 10892, NBER.

Neal, D. (2004). The measured black-white wage gap among women is too small. *Journal of Political Economy*, *112*(1), S1–S28.

Nussbaum, M. C. (2000). *Women and Human Development. The Capabilities Approach*. Cambridge, UK: Cambridge University Press.

Oswald, A. J. & Powdthavee, N. (2007). Death, happiness, and the calculation of compensatory damages. Discussion Paper 3159, IZA.

Petrongolo, B. & Olivetti, C. (2005). Unequal pay or unequal employment? a cross-country analysis of gender gaps. *mimeo*.

Pischke, J. (2007). The impact of length of the school year on student performance and earnings: Evidence from the german short school years. *The Economic Journal*, *117*(523), 1216.

Pollak, R. (2002). An intergenerational model of domestic violence. Working Paper 9099, NBER.

Pollak, R. A. (2005). Bargaining power in marriage: earnings, wage rates and household production. Working Paper 11239, NBER.

Powdthavee, N. (2009). How much does money really matter? estimating the causal effects of income on happiness. Discussion Papers in Economics 2009, The University of York.

Pudney, S. (2010). An experimental analysis of the impact of survey design on measures and models of subjective wellbeing. Understanding Society Working Paper Series 2010-01, ISER, Essex, UK.

Rao, V., Gupta, I., Lokshin, M., & Jana, S. (2003). Sex workers and the cost of safe sex: the compensating differential for condom use among calcutta prostitutes. *Journal of Development Economics*, *71*, 585.

Schafgans, M. M. & Stelcnery, M. (2006). Selectivity and the gender wage gap decomposition in the presence of a joint decision process. Research Paper No. EM/2006/513, LSE Sticerd.

Sen, A. (1990). Gender and cooperative conflicts. In I. Tinker (Ed.), *Persistent Inequalities*. New York: Oxford University Press.

Sen, A. K. (1993). Capability and well-being. In *The Quality of life* (pp.$\tilde{3}$0).

Tauchen, H. V., Witte, A. D., & Long, S. K. (1991). Domestic violence: A nonrandom affair. *International Economic Review*, *32*(2), 491.

Tella, R. D., MacCulloch, R. J., & Oswald, A. J. (2001). Preferences over inflation and unemployment: Evidence from surveys of happiness. *The American Economic Review*, *91*(1), 335.

van den Berg, B. & i Carbonell, A. F. (2007). Monetary valuation of informal care: the well-being valuation method. *Health Economics*, *16*, 1227.

van Praag, B. M. S. & Baarsma, B. E. (2001). The shadow price of aircraft noise nuisance. Discussion paper, Tinbergen Institute.

Walby, S. (2004). The cost of domestic violence. Women and equality unit, Office for National Statistics, London, UK.

Welch, F. (1979). Effects of cohort size on earnings: The baby boom babies' financial bust. *The Journal of Political Economy*, *87*, S65–S97.

Wilson, W. (1967). Correlates of avowed happiness. *Psychological Bulletin*, *67*, 294.

Winkelmann, L. & Winkelmann, R. (1998). Why are the unemployed so unhappy? evidence from panel data. *Economica*, *65*, 1.

Wu, L. L. & Li, J. C. A. (2005). Children of the nlsy79: a unique data resource. *Monthly Labor Review*, (February), 59–62.

Yang, Y., Schulhofer-Wohl, S., Fu, W. J., & Land, K. C. (2008). The intrinsic estimator for age-period-cohort analysis: What it is and how to use it. *American Journal of Sociology*, *113*(6), 1697.

Zabel, J. E. (1998). An analysis of attrition in the panel study of income dynamics and the survey of income and program participation with an application to a model of labor market behavior. *The Journal of Human Resources*, *33*(2), 479–506.