Essays on the Economics of Inequality

A thesis submitted for the degree
of Doctor of Philosophy
at the University of Leicester

by

Laura Alice Harvey

Department of Economics, School of Business
University of Leicester

January 2020

For Sheila

Abstract

Essays on the Economics of Inequality

by Laura A. Harvey

This thesis is comprised of three chapters which focus on inequality, and more closely, on the trends in inequality over time. Firstly, the second chapter addresses the issue of non-random selection into employment in the intergenerational mobility literature, by applying bounds to the distribution of wages conditional on parent income. We use the labour market attachment of the mother as a novel instrumental variable to tighten the bounds to the distribution of earnings. We find that there are substantial differences between parent income groups and changes over time by son's and daughter's. We find college to be important for all groups, but particularly for daughters. In addition, there is evidence of converging wages between sons and daughters for all parent income types over time.

The third chapter looks at the idea of a society in which everybody is the same at the same stage of the life-cycle will exhibit substantial income and wealth inequality. We use this idea to empirically quantify natural inequality - the share of observed inequality attributable to life- cycle profiles of income and wealth. We document that recent increases in inequality in developed countries are larger than observed rates would suggest. Extrapolating our measures forward suggests that natural inequalities will fluctuate over the next 20 years before settling to a new higher level.

Finally, in the fourth chapter, we document that male median real incomes have been lower than that of their forebears, at every age, for the last 30 years. We show that this is true across the life-cycle, and that younger generations have had to wait longer to reach peak earnings. Further analysis shows that this decline is particularly concentrated on high-school graduates. We further decompose the decline in labour share being more prominent for later generations, across most industries.

Acknowledgements

Firstly, I have to thank my supervisors, Dr Arek Szydlowski and Dr James Rockey for supporting me and helping me throughout this journey. From them I have learnt so much and I know I will continue to learn from them both in the future.

I would also like to thank Prof Adam Rosen and Dr Cheng Chou for their comments and suggestions which have undoubtedly improved not only my thesis but the final papers.

I will also forever be grateful and thankful to my husband for sharing this chapter of my life and being there to help me and talk through my random thoughts. I look forward to the post PhD projects we have planned.

I have met some wonderful friends and colleagues during my time at Leicester, and they will forever be the source of fond memories and support. A special thanks to Emma and Livia for all the coffee and tea breaks and help with codes!

A special mention needs to be made for my Grandma, Sheila, who has always been my number one supporter! I would also like to thank my parents for being patient with me whilst I complete my studies.

Lastly, I would like to thank everyone in the department for their support and for pushing me to do the best that I can, in addition to the financial support over the years. In particular Samantha Hill and Dr Subir Bose for all their help and advice during the past 4 years.

Declaration

Chapter 2 was produced using the High Performance Computing Facilities at the University of Leicester. It has also been presented at,

- 2018 Summer School on Socio-economic Inequality at the University of Bonn (Poster Session)

Chapter 3 is joint work with Jochen Mierau (University of Groningen) and James Rockey (University of Leicester). It is based on an earlier discussion paper Leicester 23/15 "Inequality in an Equal Society: Theory and Evidence". This paper has been presented at the following outlets:

- 2018 LIS/LWS Conference
- Seminars in Leicester, Lund, and CEPAR UNSW.

The paper corresponding to this chapter is currently under review at the time of submission of this thesis

Chapter 4 is joint work with James Rockey (University of Leicester). The paper has been presented in the following outlet,

- University of Groningen Seminar 2018

Contents

Abstract iii

Acknowledgements v

Declaration vi

List of Tables xi

List of Figures xiv

1 Introduction **1**

2 Bounds to the Distribution of Wages given Parent Income: Changes over time by Gender **4**

2.1 Introduction 4

2.2 Identification Strategy 11

2.2.1 Imposing Restrictions 13

2.3 Estimation & Inference 18

2.3.1 Inference: The Simple Case 20

2.3.2 Inference: IV Case 21

2.4 Data 23

2.4.1 Instrumental Variable in Practice 26

2.5 Results 28

2.5.1 Changes in the Distribution of Wages 28

2.5.2 Trends over Time 30

2.5.3 Within Group Inequality 33

2.5.4 Decomposing Trends 34

2.5.5 Validity and Robustness 44

2.6 Conclusion 47

3 Inequality in an Equal Society **49**

3.1 Natural Rates of Inequality 54

3.2 Inequality in an Equal Society 64

3.2.1 Inequality in the United States 65

3.2.2 Cross Sectional Time Series Analysis 68

3.2.3 Wealth Inequality 74

3.3 Inequality and the Baby Boom 77

3.4 Conclusion 84

4 The Declining Fortunes of the American Worker **85**

4.1 Introduction . . . 85

4.2 Life-Cycle Earnings . . . 92

4.2.1 Wages . . . 93

4.2.2 Hours Worked . . . 100

4.3 Individual Level Data . . . 101

4.3.1 Calculating the Labour Share . . . 103

4.3.2 Regression Analysis . . . 105

4.3.3 Explanations . . . 110

4.4 Variance Decomposition . . . 118

4.4.1 Variance Decomposition Estimators . . . 118

4.4.2 Variance Decomposition Estimates . . . 121

4.5 Hedonic Improvement . . . 127

4.6 Conclusion . . . 131

A Chapter 2 **134**

A.1 Data Appendix . . . 134

A.2 Variance of the Bounds . . . 137

A.3 Asymptotic Variance . . . 138

A.4 Results Appendix . . . 139

B Chapter 3 **144**

B.1 Proof of Proposition 1 144

B.1.1 Data Appendix 145

B.1.2 Future Cohorts Simulation Procedure 149

B.2 Additional Results 151

C Chapter 4 **158**

C.1 Data Appendix 158

C.1.1 Current Population Survey (CPS) 158

C.1.2 Summary of the Data 159

C.1.3 Economic Census 161

C.1.4 Merging the Data 163

C.2 Additional Results 165

Bibliography **173**

List of Tables

2.1 Summary Statistics 25

2.2 Bounds to the Change in the Median Wage. 32

2.3 Bounds to the Change in the Median Wage by Gender 37

2.4 Robustness Checks to the Change in Median Wage 46

3.1 Time Trends in Inequality 75

3.2 Wealth Inequality 78

4.1 Different Birth Cohorts. 93

4.2 Generational Differences in Real Wages by Industry: Unconditional Results 107

4.3 Generational Differences in Real Wages by Industry: Fixed Effect Estimates 108

4.4 Generational Differences in Real Wages by Industry: Including Covariates and Fixed Effects 109

4.5 Generational Differences in the Labour share by Industry: Including Covariates 111

4.6 Regression on Log Wage by Generation 112

4.7 Generational Differences in Real Wages by Industry: Including Covariates 113

4.8 Generational Differences in the Labour share by Industry: Unconditional Estimates 114

4.9 Generational Differences in Wage and the Labour Share: Pooled Estimates 116

4.10 Regression on Wage and Labour Share 117

4.11 Variance Decomposition 123

4.12 Variance Decomposition: Individual Breakdown 125

4.13 Variance Decomposition: For each Generation 126

A.1 Summary Statistics including SEO Sample 136

A.2 Bounds to the Change in the Interquartile Range. 139

A.3 Bounds to the Change in Interquartile Range by Gender 140

A.4 Bounds to the Change in Median Wage by Gender and Education . . 142

B.1 Country Specific Trend Estimates 155

C.1 Different Birth Cohorts. 159

C.2 Summary Statistics (CPS), Total and by Cohort 160

C.3 Industry Coding by Year in the Economic Census. 162

C.4 Summary of Excluded Industries 164

C.5 Regression on Wage by Generation 170

C.6 Regression on Labour Share by Generation 171

List of Figures

2.1 Proportion in Work by Parent Income Group and Gender over time. . 7

2.2 Probability of Working given Mothers Average Annual Hours Worked. 27

2.3 Bounds to the Median Wage 31

2.4 Change in Interquartile Range 34

2.5 Bounds to the Median Wage by Son's and Daughter's 36

2.6 Change in Interquartile Range by Gender 38

2.7 Bounds to the College Wage Premium. 39

2.8 Bounds to the Median Wage by Education Level. 41

2.9 Bounds to the Median Wage for Son's and Daughter's with College. . 42

2.10 Bounds to the change in Median Wage by Education Level and Gender 44

2.11 Distribution of Estimated Residual Wage by Labour Market Consistency 45

3.1 Income and Cohort Size by Age Group United States, 2015 57

3.2 Wealth and Cohort Size by Age Group United States, 2016 59

3.3 The Life-Cycle Adjusted Gini Coefficient 61

3.4 Actual Gini Coefficients for Labour and Total Income 67

3.5 Actual, Natural, and Excess Gini Coefficients of Labour Income for the US 1961-2015 69

3.6 Actual, Natural, and Excess Gini Coefficients of Total Income for the US 1961-2015 70

3.7 Cross Country Variation in Natural and Excess Inequality 71

3.8 Adjusted and Unadjusted Gini of Labour Income: Selected Countries: 1969-2016 72

3.9 Wealth Inequality over Time (United States) 77

3.10 The Baby Boom 79

3.11 Simulated Natural Rates of Income Inequality 81

3.12 Historical and Simulated Future Rates of Income Inequality 83

4.1 Median Wage of White Male High-school Graduates, by Decade of Birth, Over the Lifecycle 89

4.2 Median Wage by Generation over the Lifecycle 95

4.3 Median Wage (in $1000) for each Generation over time 96

4.4 Maximum Median Wage by Year Born 97

4.5 Median Wage by Generation over the Lifecycle 100

4.6 Hours Worked by Generation over the Lifecycle 102

4.7 Labour Share by Age for each Industry and Generation 105

4.8 Hedonic Improvement for Total Population (log-utility) 129

4.9 Dis-utility of Work by Year of Birth 132

A.1 Bounds to the Median Wage by Son's and Daughter's with no College. 141

A.2 Bounds to the Change in Interquartile Range by Gender and Education. 143

B.1 Income and cohort size by age group United States : 1961 151

B.2 Income and cohort size by age group United States : 2015 151

B.3 Adjusted and Unadjusted Gini of Total Income for the United States using LIS: 1974 - 2016 152

B.4 Adjusted and Unadjusted Gini of Labour Income for the United States using LIS: 1974 - 2016 153

B.5 Adjusted and Unadjusted Gini of Total Income: Selected Countries: 1969-2016 154

B.6 LIS Additional Countries, Total Income 156

B.7 LIS Additional Countries, Labour Income 157

C.1 Percentiles of the Earnings Distribution by Year 165

C.2 Maximum Median Wage (in $1000) by Year born and the Age which it was reached. 166

C.3 Maximum Median Income (in \$1000) by Year born and the Age which it was reached. 167

C.4 Median Income (in \$1000) for each Generation over time 168

C.5 Predicted Median Income (in \$1000) for each Generation by Age . . 169

C.6 Hours Worked by Generation over the Lifecycle 172

Chapter 1

Introduction

The focus of this thesis is economic inequality and its trends over time. The study of inequality, particularly of income inequality has been growing. Low levels of inequality are thought to be a hallmark of a progressive society. We look at inequality from a number of different perspectives, in addition to estimation methods.

We firstly focus on changes in intergenerational mobility, that is the importance of parental income for the income of their children. There is a substantial body of literature which tries to estimate trends in intergenerational mobility. However, it fails to appropriately take into consideration non-random selection into employment, which is especially critical given changes in the female labour market participation. As a consequence, the literature has omitted an important analysis in the trends in mobility of daughters. The second chapter reconciles these issues with partial identification methods to apply bounds to the distribution of earnings conditional on parent income, using the Panel Study of Income Dynamics (PSID). The labour market attachment of the mother is used as a novel instrumental variable to tighten the bounds to the distribution of earnings. We find that there are substantial differences between parent income groups and changes over time by son's and daughter's. We find college to be important for all groups, but particularly for daughters. In addition, there is evidence of converging wages between son's and daughters for all parent income types. However, there is evidence to suggest increasing within group inequality amongst sons which does not appear to be the case with daughters.

The third chapter looks at a selection of developed countries to test the idea that a society in which everybody is the same at the same stage of the life-cycle will exhibit substantial income and wealth inequality. We use this idea to empirically quantify natural inequality - the share of observed inequality attributable to life-cycle profiles of income and wealth. In doing so we are able to estimate levels of excess inequality - the observed in equality which cannot be explained by life-cycle income profiles. Using harmonised cross country data for both income and wealth, this chapter documents that recent increases in inequality in the United States and other developed countries are larger than observed rates would suggest. Extrapolating our measures forward, as the population pyramid returns to its long run structure following the shock of the baby boom generation, suggests that natural inequalities will fluctuate over the next 20 years before settling to a new higher level.

The focus of the fourth chapter returns to the United States. It is well documented that while real US GDP per capita has increased around 80% since 1980, median incomes have remained roughly constant. However, as the 4th chapter documents, this stagnation masks an important decline. Male median real incomes have been lower than that of their forebears, at every age, for the last 30 years. We show that this is true across the life-cycle, and that younger generations have had to wait longer to reach peak earnings. Further analysis shows that this decline is particularly concentrated on high-school graduates. The same pattern is found for female high-school graduates yet Black and Hispanic women are an important exception. While reductions in hours worked cannot explain the decline, substantial decreases in the labour share are consistent with decreasing incomes in the face of productivity growth. Calculations suggest that hedonic improvements in the quality of goods and services would have to have been 30% higher for younger cohorts consumption levels to match those of their predecessors.

Chapter 2

Bounds to the Distribution of Wages given Parent Income: Changes over time by Gender

2.1 Introduction

Concerns for limited wage growth and the equality of opportunity amongst young people has regained the attention of researchers (Bowles and Gintis, 2002), pointing to the study of the trends in intergenerational mobility, and importantly, the role of parental income in the likelihood of success for their children, and what it might mean for the levels of inequality observed in a society (see Chetty et al. (2014a,b, 2017) amongst others, and Black and Devereux (2011) for a survey). However, there is an important, under explored aspect of this literature, which is the trends in mobility for daughters and how these trends have been changing over time in comparison with son counterparts. The objective of this paper is to understand the social mobility patterns by gender and how these have changed over time using an agnostic estimation method, which reconcile issues of sample selection. We estimate bounds to the distribution of wages given parents income. By doing so, we can study the evolution of trends overall, as well as by gender and education level, conditional on parents income.

Trends in intergenerational mobility measures are important as there has been substantial changes in attitudes and policy towards achieving *equality of opportunity* and more importantly the role of women in the labour market. Such changes are important for understanding the composition of inequality in a society. Changing female labour market participation has meant that there are possible changes to intergenerational correlations, defined as the relationship between parent income and children's income, hence, raising concerns of what this might mean for the persistence of inequality as the result of family heritage. An additional contribution, is that our approach allows us to look at the effects across different points of the income distribution.

Issues in the identification of wage determining regressions for women are the result of structural changes in the women's labour market in the past hundred years (Juhn and Potter, 2006). This has meant neglect from the literature, often due to the concerns of substantial sample selection issues which arise in estimating this relationship. Hence, in order to estimate intergenerational patterns using standard methods would imply making strong assumptions over the selection mechanism. For example, to use the control function approach (Heckman, 1979) requires firstly, the correct identification of the selection equation, and secondly, that the error terms in both stages of two-stage-least-squares estimation are independent of the covariates used in each stage. Which can be a strong assumption to make and in practice might be difficult to justify.

Alternative approaches would include semiparametric selection models. They have the advantage of providing point identification without the need to make strong distributional assumptions about the selection and outcome equation. However, whilst this would give point identification, we would only be able to learn about features of the distribution with the need to impose additional restrictions such as independence or index restrictions on error terms or additional exclusion restrictions. For this reason, we opt to lose point identification but in favour of justifiable estimation assumptions.

In Figure 2.1, the proportion in work is plotted over time separately for Men and Women and by parent income group. Evidently, there has been changes in the composition of the work force over time. In particular, there are increasing levels of women in work over time, and there seems to be a relationship between them being in work and the income level of their parents. Looking at the right panel, daughters of high and middle income parents (green dashed line and blue solid line in Figure 2.1, respectively) are experiencing increased proportions in work than the daughters of low income parents (blue dotted line in Figure 2.1).

As Blundell et al. (2003) documents, changes in the composition of work force will change the observed distribution of wages, which might lead us to misinterpret changes in the actual distribution and misunderstand the importance of determining factors such as age, education and in our case, parents income, therefore, enforcing the importance of considering such selection issues when estimating wage distributions.

To reconcile issues of selection, we employ bounds to the distribution of earnings conditional on parent income. The implementation of such an agnostic estimation method allows us to firstly comment on trends in intergenerational mobility for daughters; but also to make meaningful comparisons with the trends for sons,

Figure 2.1: Proportion in Work by Parent Income Group and Gender over time.

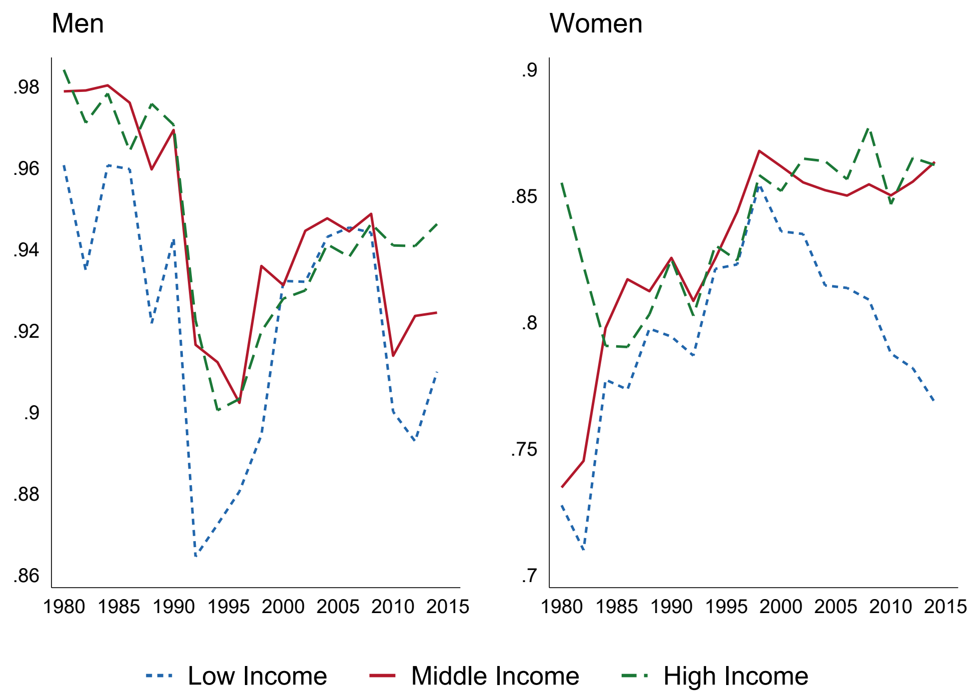


Figure Notes: On the vertical axis is the proportion of individuals in employment, and on the horizontal axis is the year. Definitions of parent income group can be found in Section 2.4.

and further looking at trends across the distribution of earnings over time. We look at two important measures: (i) between group inequality and (ii) within group inequality, which we define as changes in the distribution across parent income groups and the change in distribution within parent income classification, respectively. Additionally, we build upon so-called worst case bounds to impose restrictions motivated from economic theory and implement an instrumental variable approach. We take advantage of the comprehensive family links present in the Panel Study of Income Dynamics (PSID) and of the contained micro-data which is available for both parent and child to implement these methods.

Couch and Lillard (1998) is as far as we are aware, the first paper to document the bias caused by not controlling for sample selection in intergenerational mobility regressions. In particular, they show that estimates of intergenerational elasticity between father-sons are sensitive to the dropping of the unemployed. Alternatively, Chadwick and Solon (2002) focus on intergenerational mobility amongst daughters. Employing more traditional methods, they produce elasticities for the intergenerational mobility of daughters, and compare these for the estimates of sons, which they find to be smaller, and not always statistically significant. They take a more traditional approach to estimate their elasticities and do not consider selection into employment. More generally, the intergenerational mobility literature has focused either solely on father-son correlations or has failed to effectively take into consideration sample selection issues in any estimation strategy (Blanden et al., 2007).

There have been a limited number of papers which have used partial identification methods to measure intergenerational quantities. It is an approach which has gained a lot of momentum due to its variety of applications and the more agnostic approach to identification (see Ho and Rosen (2017) for a review of applications of partial identification methods.). There are two key papers which have used partial identification methods in the context of intergenerational mobility. Firstly, Minicozzi (2003) who considers the transmission of labour market outcomes, again, using the PSID, however restricting the sample to men. They find that estimates of intergenerational mobility are sensitive to assumptions of exogenous selection. The assumption imposed here is that the unemployed son's potential income is a function of their reason for unemployment. This allows the author to put bounds on what their income might be were they to be employed again. For example, the upper bound on earnings for someone who is looking for work would be their previous income, in another case they assume that a current student faces an upper bound income of $50,000.

Alternatively, De Haan (2011) looks at the average causal effect of parents' schooling on child's educational outcomes. They use a monotone instrumental variable as in Manski and Pepper (2000) to construct narrower bounds. Their instrument of choice is the grandparents education or the education of the other parent. They find that whilst the bounds create conservative estimates, the estimation performs well compared to other Ordinary Least Squares (OLS) and Two-Staged Least Squares (2SLS) results without the need for strong assumptions in order to achieve point identification. This work is closely related to ours, however, we extend this by looking at earnings over time and looking at differential patterns by gender and education levels.

A paper which comes close to addressing such issues is Chetty et al. (2016), however, the focus is on growing up in a disadvantaged household, and the gaps between boys and girls as adults. Our paper differs from this since we are interested in patterns across the whole parent income distribution. Moreover, very few papers have looked at patterns in intergenerational mobility measures over time. Recently, Chetty et al. (2017) looked at patterns in absolute intergenerational mobility for the United States, which they define as the proportion of children who earn more than their parents. Using historical data in conjunction with the Current Population Survey (CPS) they find that rates of absolute mobility have been declining. Due to some data limitation they find bounds to the copula of the probability of each child and parent rank pair, under various assumptions. They find falling rates of absolute income mobility over time, and a call for the growth in GDP to spread more evenly across the income distribution.

The closest paper to ours, and which is based on the same data source, is Lee and Solon (2009). They look at trends for both sons and daughters over time, trying to utilise as much of the data in the PSID as possible, in response to criticisms of other papers that disregard a lot of data and focus on elasticities solely at age 30 (An example would be Mazumder and Levine (2002)). However, Lee and Solon

(2009) fail to appropriately control for selection, particularly on the side of daughters, and favour a highly parametrised OLS model to estimate elasticities. Their findings are overall inconclusive and suggest that there hasn't been much change in intergenerational elasticities for the latter part of the twentieth century.

In this paper we estimate bounds on the distribution of earnings conditional on parent income group, using various levels of assumptions. We begin with the worst case bounds, before moving to implement further assumptions about the selection process and using an instrumental variable to tighten the bounds. Our interest lies in changes over time, which we can measure by looking at changing median wages; but also within group inequality where we take the inter-quartile range to measure. One key way in which our paper differs from the existing literature, is that our method allows us to look at how effects and trends might differ at different points in the income distribution, and to further comment on the patterns of income inequality.

Our findings suggest that whilst there has not been much change apparent within parent income group inequality, there appears to be an increasing divergence between the low parental income children and the high parental income children. Not only is this reflected in the changes in the median wage but also the returns to college, which we define as the difference in the distribution of earnings between college graduates and non-college graduates. Children of high income parents with college can expect a greater median income than their low parent income counterparts. Furthermore, we find increasing wages for all daughters over the time period we consider, whilst the counterpart sons appear to have more stagnant incomes. This convergence in sons and daughters earnings appears to have happened quicker for the children of low income parents, with the bounds on the median crossing by the end of our sample period. This suggests a driver of the increasing median incomes might be the increasing labour market participation of women. What this means in terms of inequality as a whole is that there is a divergence in the incomes of families of differing socio economic backgrounds. Simply put, the gaps between the wealthy and poor are growing. Children are further away from each other in terms of the spread of the distribution than their parents, and evidence suggests an ongoing and continuing divergence.

The paper now goes on to discuss the identification strategy in more detail, discussing each of the assumptions in turn in Section 2.2, before presenting the estimation and inference methods in Section 2.3. The data used and an explanation of the instrument is presented in Section 2.4. The results are explored in Section 2.5, including tests for the validity of our assumptions and some robustness checks in Section 2.5.5, before the Conclusion in Section 2.6.

2.2 Identification Strategy

We consider changes across the distribution of earnings and employ bounds to account for the selection bias from not observing the counter factual wage offers. To do this we analyse quantiles of the distribution to construct the inequality measures which we will be using. Thus, we will be applying bounds to the quantiles, given a set of covariates and family income.

We begin by deriving the worst case bounds (Manski, 1989). It is important to note here, that this approach is using the least number of assumptions. For example, we are not assuming linearity in covariates, or imposing a selection equation or any exclusion restrictions. The only assumption needed for estimation here is random sampling. We use this as the benchmark of our analysis, which forms the foundation of the additional restrictions we will impose.

Following the set up of Blundell et al. (2007), an individual can be either employed or not, we denote whether one is working, and thus their wage is observed, as $W = 1$, and $W = 0$ otherwise. We denote the probability of observing an individual's wage given their characteristics $X = x$, as $P(x)$, where the covariates include, for example, gender, age, education and parent income group. We are interested in looking at changes across the distribution. For this, our object of interest is the cumulative distribution function (CDF) of their wage, in which we denote Y to be the log wage, given characteristics $X = x$, which is denoted $F(y|x)$. Using the law of total probability, we can express this fully as,

$$F(y|x) = F(y|x, W = 1)P(x) + F(y|x, W = 0)[1 - P(x)] \tag{2.1}$$

Non-random sample selection means that $F(y|x, W = 0)$ is unknown and not observed in the data. This is the distribution of income observed by those who are not employed were they to chose employment. Thus, to generate our bounds we apply the definition of a cumulative distribution function which we know will be bounded between zero and one, such that $0 \leq F(y|x, W = 0) \leq 1$. Therefore, by applying this inequality equation (2.1) becomes,

$$F(y|x, W = 1)P(x) \leq F(y|x) \leq F(y|x, W = 1)P(x) + [1 - P(x)] \tag{2.2}$$

Rewriting to refer to the qth quantile of $F(y|x)$, following Manski (1994), we have the following bounds:

$$\theta_l^q(x) \leq \theta^q(x) \leq \theta_u^q(x) \tag{2.3}$$

where $\theta^q(x)$ is the qth quantile of y given x, $\theta_l^q(x)$ and $\theta_u^q(x)$ refer to the lower and upper bound of the qth quantile of y given x respectively. We can tighten these bounds by making a number of assumptions over the selection mechanism and the use of an instrumental variable. The paper now goes on to discuss these restrictions in more detail.

2.2.1 Imposing Restrictions

We combine worst case bounds with further assumptions in order to find tighter and more informative bounds, following restrictions presented in Blundell et al. (2007). We begin by imposing assumptions which are motivated by the standard idea that individuals will enter the work force if the market wage is greater than their reservation wage. These are examples of Monotone Treatment Selection (MTS) assumptions.

Stochastic Dominance

Our first restriction is a direct result of an agent's decision to select into the labour force. A higher probability of working implies tighter bounds on the distribution. This is because the width of the bounds is driven by the quantity $P(x)$. Thus, assumptions which we can make to increase this probability, taking justification from economic theory, will result in tighter bounds. Following the justification in Blundell et al. (2007), we assume that those with higher wages are more likely to enter the labour force than those with lower wages. Therefore, if we observe wages in our data, then we can state that the observed wages are first order stochastically dominated by the wages of those who did not enter the labour force. Hence, we assume:

$$F(y|x, W = 1) \leq F(y|x, W = 0) \qquad \forall y, \forall x \tag{2.4}$$

for each y with $0 < F(y|x) < 1$. Similarly, this relationship can be written as in Equation 2.5.

$$Pr(W = 1|Y \leq y, x) \leq Pr(W = 1|Y > y, x) \tag{2.5}$$

If we substitute the assumption from equation (2.4) into equation (2.1), this implies that the bounds on the distribution of wages will then become,

$$F(y|x, W = 1) \leq F(y|x) \leq F(y|x, W = 1)P(x) + [1 - P(x)] \tag{2.6}$$

The Stochastic Dominance assumption may fail under certain circumstances, for example if, for some groups, there is a strong positive relationship between wage and reservation wage, this might be bought about by not considering the role of assets in the reservation wage and labour market wage relationship. Thus a sufficient condition which is required, under which the Stochastic dominance restriction will hold is that potential labour income is conditionally independent of reservation wage given the individuals characteristics.

Median Restriction

Given the strength of the Stochastic Dominance restriction, a weaker version of this assumption can be imposed, by arguing that if an agent has earnings above the median wage, they are more likely to enter the work force than those that do not have earnings above the median wage. This implies the following inequality would hold for all individuals with wages above the median wage,

$$0.5 \leq F(y|x, W = 0) \leq 1, \quad \text{if} \quad y \geq y^{50(W=1)}(x)$$

where $y^{50(W=1)}(x)$ denotes the median wage of the those whose wages we observe conditional on x. In this case, there would be tighter bounds on the distribution above the median value such that our bounds become,

If $y < y^{50(W=1)}$

$$F(y|x, W = 1)P(x) \leq F(y|x) \leq F(y|x, W = 1)P(x) + [1 - P(x)] \tag{2.7}$$

If $y \geq y^{50(W=1)}$

$$F(y|x, W = 1)P(x) + 0.5[1 - P(x)] \leq F(y|x) \leq F(y|x, W = 1)P(x) + [1 - P(x)]$$

In summary, the stochastic dominance and median restrictions formally present the ideas that, *ceteris paribus*, if you have a higher offered wage you will be more likely to work. These assumptions are imposing positive selection into the work force. We formally test this and cannot reject these assumptions, as presented in Section 2.5.5. Our final assumption, which is in the form of an instrumental variable restriction is presented in more detail in the following sub-section.

Instrumental Variable

We can further sharpen our bounds by using Monotone Instrumental Variable (MIV) methods proposed by Manski and Pepper (2000). The IV assumption does not place any restrictions over selection into the labour market, unlike the previous Median and Stochastic Dominance restrictions. We begin by exploring the stronger exclusion restriction from Manski (1994). To fix ideas, suppose we have an instrument Z, which has an effect on the individual's labour market participation decision but does not affect our outcome of interest, that in our case is their compensation from labour. Under this exclusion restriction we have that,

$$F(y|x, z) = F(y|x)$$

Following Manski (1994)[1] under this assumption, the worst case bounds become,

$$\max_z\{F(y|x, z, W = 1)P(x, z)\} \leq F(y|x) \leq \min_z\{F(y|x, z, W = 1)P(x, z)+[1-P(x, z)]\} \tag{2.8}$$

It is evident that we can combine the bounds in equation (2.8) with the previous median and stochastic dominance restriction to further tighten the bounds. This is done by replacing the lower bound with the lower bound in equation (2.4) for Stochastic Dominance or lower bound in equation (2.7) for the Median Restriction.

Monotone Instrumental Variable

Additionally, we can weaken the exclusion restriction to the monotone instrumental variable restriction as in Manski and Pepper (2000) and Manski and Pepper (2009), if one believes that the exclusion might not be credible. Thus, we can assume the direction of the relationship; that the distribution of wages is monotonically increasing in the value of the instrument. Such that,

[1] This is the case where Y is independent of Z conditional on X.

$$F(y|x,z') \leq F(y|x,z) \qquad \forall y, x, z, z' \text{ with } z < z'$$

This suggests that as the value of the instrument, Z, increases, for a given value, the distribution of incomes will stochastically dominate that with a lower value of Z.

The bounds are tighter under the MIV as the tightest bound given the value of Z is found and then integrate out Z.

Formally, for each $F(y|x,z)$, the best case lower bound is the value which is highest among those on the support of Z with $z \leq z_1$, then we can rewrite the lower bound for a value $Z = z_1$ as,

$$F(y|x,z_1) \geq F^l(y|x,z_1) \equiv \max_{z \leq z_1}\{F(y|x,z,W=1)P(x,z)\} \tag{2.9}$$

Similarly, for each $F(y|x,z)$, the best upper bound, minimises the value among those on the support of Z with $z \geq z_1$, such that,

$$F(y|x,z_1) \leq F^u(y|x,z_1) \equiv \min_{z \geq z_1}\{F(y|x,z,W=1)P(x,z) + [1 - P(x,z)]\} \tag{2.10}$$

Therefore the bounds to the conditional distribution of wages where we have monotone instrumental variable becomes,

$$E_Z[F^l(y|x,Z)|x] \leq F(y|x) \leq E_z[F^u(y|x,Z)] \tag{2.11}$$

By integrating over the distribution of Z given $X = x$ gives the bounds to the distribution $F(y|x)$.

The intuition behind these bounds comes from the exogenous variation in the instrument, this provides us with tighter bounds given sufficient variation in the instrumental variable. We prefer the weaker monotonicity assumption to exclusion restrictions, thus our findings which use the IV refer to the monotonicity assumption.

In the case of the MIV estimation we cannot employ the same inference methods as with the worst case bounds, stochastic dominance and median restriction cases. In those cases the intervals for $F(y|x)$ are estimated using asymptotically normal estimators. As a result, the confidence intervals are simple enough to compute. However in the MIV case, we have the case of conditional moment inequalities and so an alternative inference method is required. These various methods are discussed in the next section.

2.3 Estimation & Inference

We estimate the conditional distribution non-parametrically, and conduct inference using the approach of Imbens and Manski (2004) for the simple case and use Andrews and Shi (2013) for the MIV case. To take our cumulative distribution functions to the quantiles in order to calculate our measures for inequality, we use Manski (1994), which involves the calculation of various components. Firstly, the probability of working, $P(x)$, from equation (2.2) we estimate to be,

$$\hat{P}(x) = \frac{\sum_{i=1}^{N} \mathbb{1}(W_i = 1)\gamma_k(x_i)}{\sum_{i=1}^{N} \gamma_k(x_i)} \tag{2.12}$$

where we define weights, $\gamma_k(x_i)$, which are for each group k to be

$$\gamma_k(x_i) = \mathbb{1}(year_i = year_k)\mathbb{1}(educ_i = educ_k)\mathbb{1}(gender_i = gender_k)$$
$$\mathbb{1}(age_i = age_k)\mathbb{1}(parentinc_i = parentinc_k)$$

Thus, the probability of working is just the average number of individuals working in a group of similar characteristics in terms of year of survey, education level, gender, age and parent income group, as defined by the weight $\gamma_k(x_i)$ which is the indicator of the numerous x covariates.

To derive the estimates for the conditional distribution of wages, we use the following kernel estimator,

$$\hat{F}(y|W_i = 1, x_k) = \frac{\sum_{i=1}^{N} \Phi((y - y_i)/h)\mathbb{1}(W_i = 1)\gamma_k(x_i)}{\sum_{i=1}^{N} \mathbb{1}(W_i = 1)\gamma_k(x_i)} \tag{2.13}$$

where Φ is the standard normal CDF. Note, that we smooth the distribution in order to allow for the unique mapping of the quantiles to earnings. Parameter h is the bandwidth which we fix to one-fifth the standard deviation of wages in each group, following the procedure which is used in Blundell et al. (2007).

To limit the number of groups, we consider three broad categories of age: young, middle and old. Furthermore, we group education of those with and without some college, in addition to pooling years. This is in order to ease the estimation procedure and limit any empty cells.

Our estimation approach will change slightly to allow for the instrumental variable, as now we need to include our instrument z in our calculation of $P(x)$, the probability of working. We redefine our weights, now $\gamma_k(x_i, z_i)$ which include the instrument z_i.[2] Hence our weights become,

$$\gamma_k(x_i, z_i) = \mathbb{1}(year_i = year_k)\mathbb{1}(educ_i = educ_k)\mathbb{1}(gender_i = gender_k)$$
$$\mathbb{1}(age_i = age_k)\mathbb{1}(parentinc_i = parentinc_k)\mathbb{1}(z_i = z_k)$$

The estimation of the conditional wage distribution is still the same as discussed in equation (2.13).

It is important to note that we have an estimation limitation in that we can only present non-trivial bounds for a quantile q where $q \geq 1 - P(x)$ (for lower bound estimates) and $q \leq P(x)$ (for upper bound estimates) this is because at these points the bound will be $\pm$ infinity. Recalling from the previous section, the width of the bounds is determined by the probability of work $P(x)$, so at the extremes of the distribution there might only be one known bound for a given quantile, as the other is bounded at $\pm$ infinity. We now move on to discuss how we conduct inference in this framework.

2.3.1 Inference: The Simple Case

As previously mentioned we need to employ two cases to deal with inference. Firstly, we can employ confidence sets such as those set out in Imbens and Manski (2004), for the case where we have worst case bounds, stochastic dominance and median restriction assumptions. Let $\hat{\theta}_i$ denote the estimate for either the upper or lower bound.[3] Then the confidence sets will take the following form,

$$CI_{\alpha}^{\theta} = \left[\hat{\theta}_l - C_N \cdot \frac{\hat{\sigma}_l}{\sqrt{N}}, \hat{\theta}_u + C_N \cdot \frac{\hat{\sigma}_u}{\sqrt{N}}\right] \tag{2.14}$$

[2] This is discussed further in Section 2.4.1.

[3] For ease of notation, we drop the superscript q in the discussion of inference.

Where $\hat{\sigma}_l$ and $\hat{\sigma}_u$ refer the standard deviations of the lower and upper bound of the bounded distribution respectively. The derivations of the standard deviations can be found in Appendix A.2. The quantity, C_N, must satisfy the following condition,

$$\Phi\left(C_N + \sqrt{N} \cdot \frac{\hat{\Delta}}{max(\hat{\sigma}_l, \hat{\sigma}_u)}\right) - \Phi(-C_N) = \alpha \tag{2.15}$$

Where we have that $\hat{\Delta} = \hat{\theta}_u - \hat{\theta}_l$, the difference between the lower and upper bound estimates[4], and α refers the significance level.

2.3.2 Inference: IV Case

We can not use Imbens and Manski (2004) to construct confidence intervals in the case of MIV because for this class of bounds the asymptotic distribution of the estimates is complicated and difficult to approximate. A number of papers have tackled this issue in different ways, however we proceed to use Andrews and Shi (2013) to construct confidence intervals in the MIV case[5].

From section 2.2.1, we know that with the introduction of a monotone instrumental variable, denoted z, means that the bounds can be written as,

$$\sup_{z \leq z_1} \theta_l(x, z) \leq F(y|x, z) \leq \inf_{z \geq z_1} \theta_u(x, z) \tag{2.16}$$

This can be represented by the following moment inequalities,

[4]By Lemma 3 of Stoye (2009) super-efficiency is implied in this case.

[5]Notably Chernozhukov et al. (2013) and Lee et al. (2013) would be the main alternative to Andrews and Shi (2013), which is the approach that we choose. There is no difference in how each of these approaches perform. We choose Andrews and Shi (2013) due to the availability of code and ease of implementation (see Andrews et al. (2017)). For a survey of inference methods in partially identified models see Canay and Shaikh (2017).

$$F(y|x) - F(y|x, z, W = 1)P(x, z) \geq 0 \ \ \forall z \tag{2.17}$$

$$F(y|x, z, W = 1)P(x, z) + 1 - P(x, z) - F(y|x) \geq 0 \ \ \forall z$$

Following inference procedures outlined in Andrews and Shi (2013), as before, we want the confidence intervals for the true parameter, θ_0 (in the above notation this is $F(y|x)$), however in our case this is not point identified. Yet, we want the nominal coverage of $1 - \alpha$ for $\alpha \in (0, 1)$ for θ_0.

The construction of confidence intervals is based on the inversion of a test statistic to test the null $H_0 : \theta_0 = \theta$ where the parameter of interested might be set identified by conditional moment inequalities. The test is of the standard form,

$$\phi_n(\theta) = \mathbb{1}\{T_n(\theta) > c_n(\theta, 1 - \alpha)\}, \tag{2.18}$$

where α is the nominal level of the test, $c_n(\theta, 1 - \alpha)$ is a simulated critical value and $T_n(\theta)$ is the test statistic. This test can then be inverted and used to construct confidence intervals for the parameter of interest. Thus, confidence interval takes the form,

$$CI_n(1 - \alpha) = \{\theta \in \Theta : \phi_n(\theta) = 0\} \tag{2.19}$$

The main idea behind this approach is to construct the test statistic based on unconditional moment inequalities/equalities which have been transformed from the initial conditional moment inequalities/equalities by some weight, $g \in G$, from the instrument. From the unconditional moment inequalities, a sample average and sample variance is computed to be used in the test statistic. Specifically, the null hypothesis is rejected for large values of the test statistic, $T_n(\theta)$,

$$T_n(\theta) = \int T_n(\theta, g) dQ(g) \tag{2.20}$$

where $T_n(\theta, g)$ is the test statistic of Cramér-von Mises form, from the transformed moment inequalities and Q is the weight function for $g \in G$. Additionally, critical values are approximated based on generalised moment selection (Andrews and Soares, 2010) and estimated used Gaussian asymptotic approximation.[6] [7]

The paper now goes on to discuss the data used in the next section and justification for the choice of monotone instrumental variable, before finally the results.

2.4 Data

We take advantage of the parent-child links in the Panel Study of Income Dynamics (PSID), a survey which ran annually from $1969 - 1997$ and then alternate years from $1997 - 2015$. The PSID interviewed and followed a representative sample of U.S. households, such that any splits were followed and, more importantly for our purposes, their children were followed into adulthood. For some families, three generations are captured in the PSID. As a result of this design, the survey contains detailed micro data on different generations of the same family such as earnings and other demographic and education variables for children and their parents.

[6] Other parameters are set to the default as discussed in Andrews et al. (2017).

[7] To apply this method to calculate the 90% confidence interval to the change over time, we take the difference in the confidence intervals around the bounds which are at the 5% level. Therefore the confidence intervals for the bounds to the change over time will be conservative, containing the parameter of interest with probability at least 0.9 asymptotically.

There are a number of supplements in the PSID. Notably, we drop from our core sample anyone from the survey of economic opportunity, a supplement of the PSID which over sampled poorer households. The PSID is discussed in more detail in Appendix A.1[8].

The outcome of interested will be the cumulative distribution function of children's labour income, that is solely their earnings from employment. To limit the number of groups for which we need estimate distributions for, we create a number of discrete variables for our controls. Firstly, we consider two education groups; those with high school or less and those with at least some college. We expand this further to include, has a college degree or not, and a dummy for graduating high school. Secondly, we group age into young ($18 \leq age \leq 35$), middle ($35 < age \leq 50$) and older ($50 \leq age \leq 65$), dropping anyone who cannot be classified into these groups. Our findings are robust to different classifications of our controls. An additional measure we use is to pool years in the pre 1997 data.

We group the children in the sample into three classifications, using the average earnings for the parents during their time in the PSID to act as a proxy for their life time earnings. Firstly, our low income group which have an average parents income in the bottom quantile, and children with parents income in the top quantile we consider as our high income group. Lastly, we consider a medium income group which is children whose parents were between the $25th$ and $75th$ quantile.[9] This variable in conjunction with the above discussed groups: Age, education and year form our conditioning variables which we refer to in the vector x. Thus all of our restrictions are being made conditional on these variables. Thus, to ensure we have sufficient observations for children, we limit the analysis to the trends post 1980.

[8] Our findings are robust to the inclusion of the survey of economic opportunity. In Section 2.5.5 we present results which include the SEO sample and find minimal differences.

[9] The qualitative results are robust to the changing of the cut off in the parent classification. See section 2.5.5 where we present some results that alter this definition here. We find estimates broadly in line with our main findings.

We include those individuals where we are able to observe their parents when they were a child (less than eighteen years of age).

Table 2.1: Summary Statistics

	Overall		***Low Income***		***Middle Income***		***High Income***	
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
Children								
Female	0.53	0.50	0.57	0.50	0.52	0.50	0.52	0.50
Age	35.55	7.71	35.26	7.62	35.84	7.75	35.28	7.68
White	0.90	0.30	0.75	0.43	0.94	0.24	0.97	0.17
Black	0.08	0.27	0.21	0.41	0.05	0.21	0.01	0.10
Hispanic	0.01	0.09	0.02	0.14	0.01	0.08	0.00	0.05
Other Ethnicity	0.01	0.12	0.02	0.14	0.01	0.11	0.02	0.12
Years of Education	13.66	2.10	12.61	1.99	13.64	1.98	14.75	1.89
High School Graduate	0.92	0.27	0.83	0.37	0.95	0.23	0.97	0.17
College Graduate	0.31	0.46	0.14	0.35	0.29	0.45	0.53	0.50
In Employment	0.83	0.37	0.79	0.41	0.84	0.36	0.85	0.36
Labour Income	32,575	49,603	22,152	20,791	32,322	51,492	43,428	62,202
Mother								
White	0.52	0.50	0.40	0.49	0.54	0.50	0.60	0.49
Black	0.05	0.22	0.14	0.34	0.03	0.17	0.01	0.09
Hispanic	0.00	0.07	0.01	0.11	0.00	0.05	0.00	0.01
Other Ethnicity	0.42	0.49	0.45	0.50	0.43	0.49	0.39	0.49
High School Graduate	0.74	0.44	0.46	0.50	0.78	0.41	0.92	0.27
College Graduate	0.13	0.34	0.04	0.20	0.09	0.28	0.30	0.46
Average Hours Mother Worked per Year	986	814	890	807	1,014	793	1,020	853
Father								
White	0.48	0.50	0.29	0.46	0.51	0.50	0.59	0.49
Black	0.03	0.17	0.07	0.26	0.02	0.14	0.01	0.09
Hispanic	0.01	0.08	0.01	0.12	0.00	0.06	0.01	0.08
Other Ethnicity	0.48	0.50	0.62	0.49	0.46	0.50	0.39	0.49
High School Graduate	0.72	0.45	0.41	0.49	0.72	0.45	0.92	0.27
College Graduate	0.24	0.43	0.04	0.18	0.17	0.37	0.51	0.50
Average Hours Father Worked per Year	2,196	693	1,831	952	2,199	621	2,416	502
Family								
Average Family Income	61,887	44,698	24,651	8,117	55,304	10,841	112,243	60,628
Observations	81,812		20,458		40,882		20,472	

Note: This table is produced excluding those in the SEO sample. Parent averages come from the average of income or hours worked when the child was aged between 14 and 18. All monetary variables are in terms of 1999 USD.

In table 2.1, the mean and standard deviations are presented for the overall sample and by parent income classification group. The first thing to note is regarding demographics, which one wouldn't expect to change depending on parent income group, do not vary by their parent classification. We have an even male-female split for the total sample and for the income sub samples. Additionally, as one might expect, education is increasing in parent income, both in terms of high school graduation rates and college degree attainment. In terms of parents, the average hours mothers worked has substantial variation, where low income mothers work only 890 on average compared to the high income counterparts who work 1,020 on average. This variation in hours worked is going to form the basis of the instrumental variable. We now present our choice of IV and its justification, before moving on to present our findings.

2.4.1 Instrumental Variable in Practice

Following our discussion in Section 2.2.1, we can use our instrument, z, as either an excluded or monotone instrument. Under the exclusion restriction, the assumption is that the distribution of wages, y, and the instrument, z, are independent conditional on covariates, x. In other words the distribution of wages does not vary with the value of the instrument. The monotone instrumental variable restriction imposes a weaker version of this assumption, that is that the conditional distribution of wages is weakly increasing in the value of the instrument.

A number of papers have documented possible intergenerational correlations in attitudes to work. Toledo (2007) looks at intergenerational transmission of work hours from father to sons. The paper documents that if the father works more than the average for their cohort then it is likely that their son will also. Additionally, Fernández et al. (2004) argue that the increases in female labour market participation over the past century are the results of individuals growing up in a household with a working mother. Using the need for women to work in the war-era, they argue this formed a new type of family dynamic; one in which the mother worked. As a result this led to changing attitudes towards female employment and consequentially, increased their labour market participation. With this evidence, we take a measure of the strength of the labour market attachment of the mother to be our instrument. That is, we take the average number of hours that the mother worked per year when the individual was younger (between the ages of 14 and 18), thus fully utilising the panel nature of the PSID.

By controlling for parental income, mothers' labour market attachment should not directly affect their children's wages. Even if it does, this would imply that more attachment should mean higher wages. So at the very least the Monotone Instrumental Variable assumptions should be satisfied.

Figure 2.2: Probability of Working given Mothers Average Annual Hours Worked.

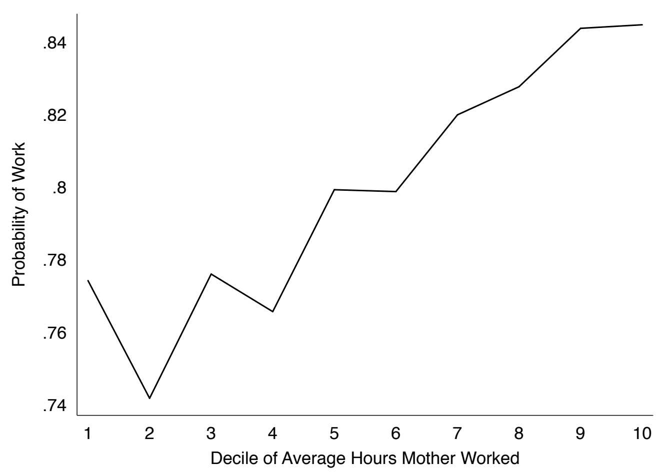


Figure Notes: This used the entire PSID sample, where the probability of working is the average of those working given the decile of their mothers working hours during their teenage years.

In Figure 2.2, the relationship between the probability of work and the decile of mothers average annual working hours shows a positive relationship.[10] That is, that the child's probability of entering the labour market is increasing in how much their mother worked. The bottom decile of hours worked for mother implies the child

[10]This is based on the total sample of the PSID. The probability of working is the average percentage which were working given the decile their mothers working hours fell into.

has around a 75% probability of entering the workforce, conversely at the top of the mothers labour market attachment is almost a 85% likelihood of working.

2.5 Results

Here we present the main findings under various levels of assumptions. Firstly, we focus on trends over time, in terms of within group and between group inequalities. We then develop this further by decomposing trends by gender and education levels. Firstly, we define formally how we measure these changes and levels of inequality.

2.5.1 Changes in the Distribution of Wages

As discussed earlier, we can think about the trends in two ways: within and between group inequality. We firstly consider between group inequality, which we define as the inequalities and differences that exist between parental income groups. We use changes in the median wage as our primary measure of between group inequality such that,

$$\Delta\theta^{q=0.5} = \theta^{q=0.5}_{i,t'} - \theta^{q=0.5}_{-i,t}$$

Where i refers to either the upper or lower bound and $-i$ is the other bound. Additionally, t', t are different points in time with $t' > t$. It easy to see how this can be generalised to look at changes of different quantiles.

We are interested in how there have been movements not only between groups but also within parental income groups. To measure inequality within group we are going to use the interquartile range (IQR), and we want to apply bounds to this measure, thus we take the difference between the upper and lower bounds to the change and to also bound this change over time. For example the upper bound on the change in IQR will be the lower bound on the $75th$ quantile minus the upper bound on the $25th$ quantile. See this formally below,

$$IQR_u = \theta_l^{q=0.75} - \theta_u^{q=0.25} \tag{2.21}$$

$$IQR_l = \theta_u^{q=0.75} - \theta_l^{q=0.25}$$

$$\Delta IQR_i = IQR_{i,t'} - IQR_{-i,t}$$

For example, we might have $t' = 2014$ and $t = 1980$, or any other time window of interest.

In addition to this we also want to consider bounds to wage differentials and their change over our time period by gender, education and parental income group. Suppose we want to bound the wage differential by some characteristic x, we want to know ${D^q}_t = \theta_t^q(x_1) - \theta_t^q(x_0)$. This could be for example the difference in wages at time, t by characteristic x, which could, for example, be gender. Therefore its evident that the bounds to this differential will be,

$$\theta_{l,t}^q(x_1) - \theta_{u,t}^q(x_0) \leq {D^q}_t \leq \theta_{u,t}^q(x_1) - \theta_{l,t}^q(x_0) \tag{2.22}$$

This measure is presented in relation to the returns to college education when we decompose trends further. This paper now goes on to formally present its findings.

2.5.2 Trends over Time

Firstly, we look at the evolution of the median wage as our measure of between group inequality between children of different parent income groups. Figure 2.3 plots the upper and lower bound to the median wage from 1980 to 2014 for our three parent income groups, along with the 90% confidence interval under the median restriction assumption.

What's striking, is the dominance between the median wage of the high and middle income parent groups over that of the low income parent group. Over the time period considered, the lower bound median wage for the high group remains between 10 and 10.5 in contrast with the low parent income group, where the lower bound median wage remains between 9.5 and 10. Furthermore the lower bound median wage for middle income parents exceeds the upper bound for the low income parents in each year.

Secondly, there is an apparent upward trend for all parent income groups, as demonstrated in the lower bound in 2014 being greater than the upper bound in 1980 for the low and high parent income group in particular. This growth does not appear to be as strong for the children of middle income parents. We look at this apparent growth in median wage in more detail.

In table 2.2 the changes in the median wage from 1980 to 2014 are presented along with 90% confidence intervals, for different sets of assumptions. One thing to note is that the Median Restriction is nested in the Worst Case bounds and similarly the Stochastic Dominance restriction is nested in the Median Restriction by construction. The IV is nested solely within the worst case, but we deem this assumption to be a compromise between this and the stochastic dominance assumption.

Figure 2.3: Bounds to the Median Wage

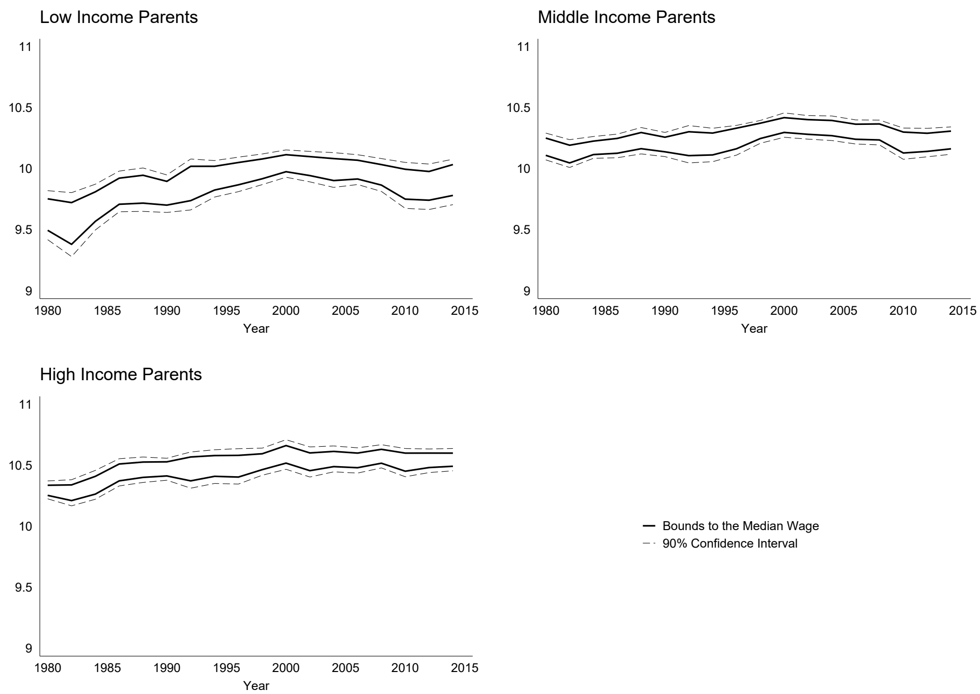


Figure Notes: On the vertical axis is the log median annual wage by year on the horizontal axis. Estimates are produced with the median restriction assumption along with the 90% Confidence intervals.

Returning to the discussion of the results, stikingly the high income parent group is the only group with positive bounds to the change, given all assumptions, along with statistically significant change over the time period. This suggests that they are the only group that we can confidently say have experienced a growth in their median wage. The change in the worst case for high income parents has a lower bound change of 0.09 log points, which is substantial compared to what is reported for the worst case of all other groups. For other parent income groups, in the worst case, experience negative median wage growth conditional on parent income. However imposing monotonicity assumptions appears to significantly sharpen the bounds on the change. Under this assumption individuals of all parent income groups have experienced growth in their median wage, however the high income group appears to have experienced those most at a change of 0.25 log points compared to 0.06 and -0.08 log points for the low and middle income groups, respectfully.

Table 2.2: Bounds to the Change in the Median Wage.

	Low Income	Middle Income	High Income
Worst Case	-0.17, 0.67 [-0.19, 0.69]	-0.19, 0.27 [-0.21, 0.29]	0.09, 0.41 [0.07, 0.43]
Median Restriction	0.03, 0.54 [0.00, 0.56]	-0.09, 0.20 [-0.11, 0.22]	0.16, 0.35 [0.14, 0.37]
Monotonicity (IV)	0.06, 0.43 [-0.47, 1.03]	-0.08, 0.22 [-0.39, 0.54]	0.25, 0.36 [-0.15, 0.94]
Stochastic Dominance	0.11, 0.46 [0.09, 0.48]	-0.02, 0.15 [-0.04, 0.16]	0.19, 0.31 [0.17,0.33]

Note: The 90% Confidence Interval for the change is given in square brackets

There is also evidence to suggest we cannot reject the hypothesis that the children of the middle income group have not experienced growth in their median wage over time. The lower bound almost always is below zero and changes are not statistically significant. This evidence suggests that children of high income parents are moving further away from children of low income parents in the distribution. This additional dispersion will add to high levels of inequality observed in a society. Moreover, if this is the product of intergenerational transmissions, it can be expected that the spread in the distribution can be expected to get worse if this pattern continues, leading again to higher levels of observed inequality. However, it is important to note that most of the growth in wages of the high income children occurred prior to the turn of the century, with more stable median wages being seen since 1996.

2.5.3 Within Group Inequality

Having looked at broad trends in inequality between children of different parent income groups, our focus now shifts to looking at the dispersion of wages within parent income groups, and how this has changed over time as measured by the Interquartile Range as discussed in equation (2.21).

Figure 2.4 presents the change in the interquartile range between 1980 and 2014 under the Median Restriction assumption and parent income group. The limits of the box show the upper and lower bound to the change and the whiskers refer to the 90% confidence interval.

It is immediately evident that none of the changes are statistically significant at the 10% level for any income group; meaning that we can not reject that there hasn't been minimal changes within parent income group levels of inequality. Thus, these results do not carry evidence that the distributions of children's wages of different parent groups are more disperse. In summary, supporting the suggestion that children of high income parents are moving further away from the children of low income parents across the whole distribution, yet not growing further apart within parent income group, hence, contributing to the levels of inequality observed.

Figure 2.4: Change in Interquartile Range

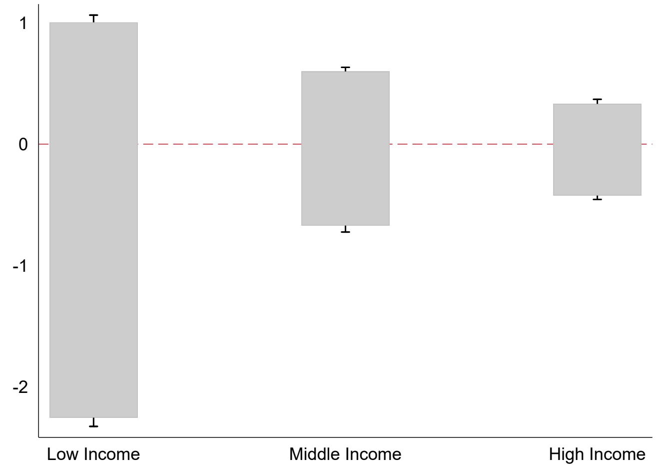


Figure Notes: Change in the IQR under the Median Restriction. The limits of the box show the upper and lower bound to the change and the whiskers refer to the 90% confidence interval. The results under the Worst Case assumptions and stronger Stochastic Dominance assumptions are presented in Table A.2 in Appendix A.4.

2.5.4 Decomposing Trends

Moving on to further understand the trends over time of within and between group inequality, we decompose further on two dimensions. By gender and by education level. An important advantage to our approach is the ability to make meaningful comparisons over time for both Men and Women, a feature which has not been as explored in previous studies. In addition to this we can explore the inter-play between gender and education level. We split education groups by those with some college and those with no college education. Here, selection will play a key role, as this might have changed over time for sons and daughters, in addition to the mechanism behind such selection into employment.

Differences between Sons and Daughters

Looking at patterns in sons' and daughters' wages over time, Figure 2.5 plots the median wage by parent income group and gender under the median restriction along with the 90% confidence intervals for the period 1980 to 2014. Decomposing changes in the median wage for both sons and daughters firstly leads to an unsurprising result. The median wage of sons for each parental income group is higher than that of their female counterpart. This dominance is persistent throughout the time period studied.

Not only is there dominance between sons and daughters, but there also appears to be dominance across parent income groups for all children. In 2014, bounds to sons' median wage for the high income group were focused about 11.5, and for daughters the bounds were between 10 and 10.5. In contrast, the bounds to the median wage for the children of low income parents were between 9 and 10 for daughters and around 10 for son's. Showing that, regardless of gender, there appears to be a benefit of being from a high income family.

Additionally, one can see that there is evidence of convergence in the median wage between sons and daughters in each parent income category. This convergence seems to mostly be driven by median wage growth for daughters. Informally, it seems the wages of daughters are catching up with those of the equivalent son's over time. Possibly this is evidence of the impact of efforts to close the gender wage gap. Strikingly, the difference seems higher for the high parent income group than for our other parent income categories, as son's wages appear to have minimal to no median wage growth.

To look at the growth in the median wage in more detail, we can estimate bounds to the change in the median wage for sons and daughters and the 90% confidence

Figure 2.5: Bounds to the Median Wage by Son's and Daughter's

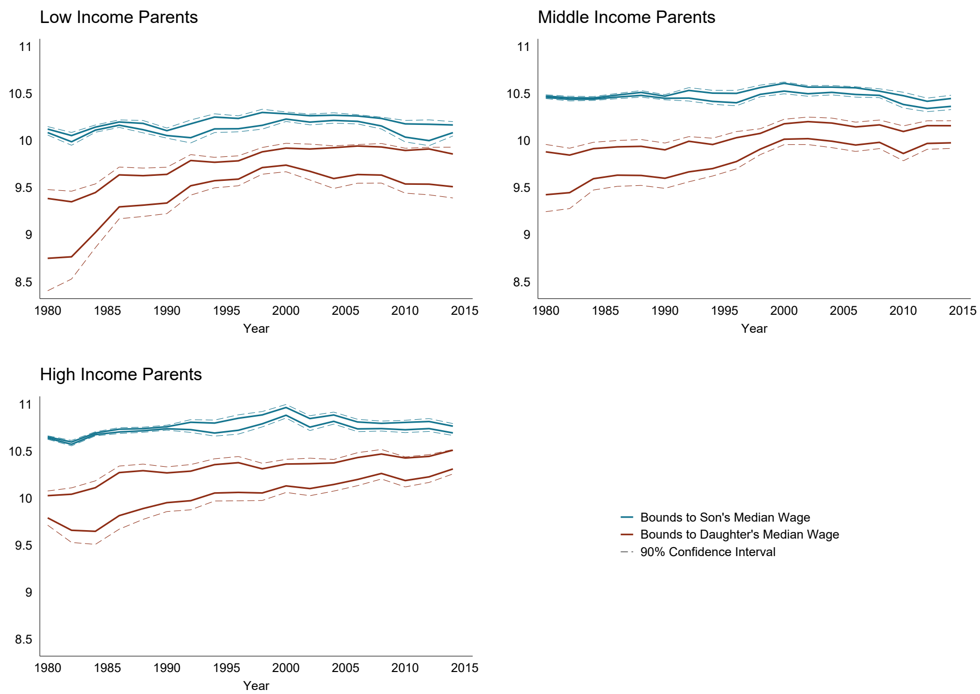


Figure Notes: On the vertical axis is the log median annual wage and year on the horizontal axis. Estimates are produced with the median restriction assumption along with the 90% Confidence intervals.

interval for each of our assumptions as presented in Table 2.3. Firstly, note that only the high income group experienced positive median income growth under all of our assumptions. However, in line with previous thoughts the daughters of high income parents experienced the biggest growth in their median wage. Under the Median Restriction, they experienced a lower bound change of 0.28 log point, significantly higher than the equivalent sons who experienced only a lower bound change of 0.04 log points.

Looking at the low and middle income parent income groups, we cannot conclude whether son's experienced any income growth. Under the stronger assumptions of Median Restriction and Stochastic Dominance, the middle income sons appeared to experience a decline in their median wage over the time period considered. Al-

Table 2.3: Bounds to the Change in the Median Wage by Gender

	Low Income	Middle Income	High Income
Son's			
Worst Case	-0.07, 0.16	-0.12, 0.04	0.04, 0.14
	[-0.09, 0.17]	[-0.13, 0.05]	[0.03, 0.15]
Median Restriction	-0.04, 0.08	-0.11, -0.01	0.04, 0.13
	[-0.05, 0.10]	[-0.12, 0.00]	[0.03, 0.14]
Monotonicity (IV)	-0.04, 0.08	-0.10, 0.01	0.07, 0.10
	[-0.62, 0.54]	[-0.31, 0.04]	[0.04, 0.81]
Stochastic Dominance	-0.04, 0.07	-0.11, -0.03	0.05, 0.10
	[-0.05, 0.09]	[-0.12, -0.02]	[0.03, 0.11]
Daughter's			
Worst Case	-0.21, 1.34	-0.13, 0.84	0.21, 0.77
	[-0.25, 1.37]	[-0.15, 0.87]	[0.19, 0.78]
Median Restriction	0.13, 1.11	0.10, 0.73	0.28, 0.72
	[0.08, 1.13]	[0.07, 0.76]	[0.26, 0.74]
Monotonicity (IV)	0.11, 0.66	0.06, 0.61	0.40, 0.58
	[-0.02, 1.67]	[-0.58, 1.20]	[0.02, 0.99]
Stochastic Dominance	0.17, 1.06	0.18, 0.68	0.36, 0.64
	[0.13, 1.09]	[0.15, 0.70]	[0.34, 0.65]

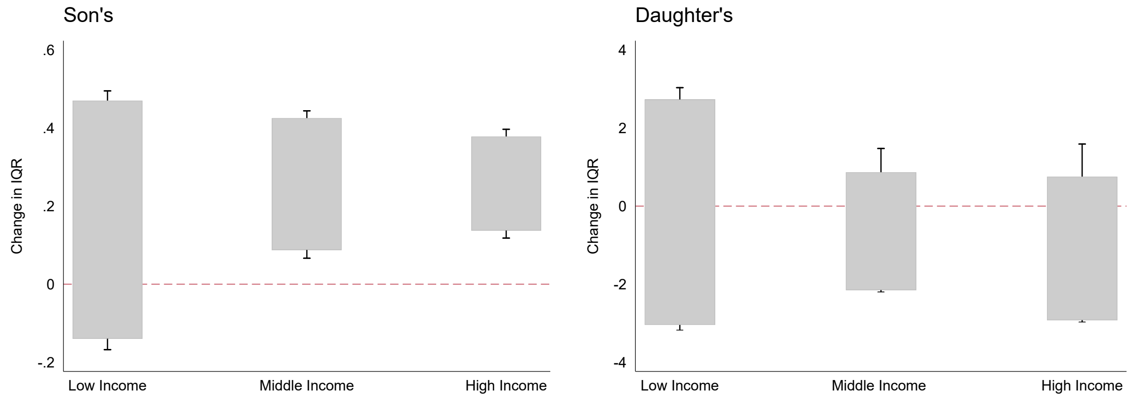
Note: The 90% Confidence Interval for the change is given in square brackets

ternatively, under stronger assumptions for daughters we can conclude that they did experience positive median wage growth, with the children of low income parents seeing a 0.13 log point increase and 0.10 log point increase for middle income, again under the conservative Median Restriction. This suggests that growth in median wages is mainly being driven by the growth in the median wage of high income parent's children, and in particular the growth of daughter's median wage.

Additionally, we can look at how the distribution has changed within son and daughter groups. Figure 2.6 plots the change in the interquartile range between 1984 and 2014,[11] panel (a) shows the change for sons given parent income group,

[11]This is to limit missing values from undefined bounds at the extreme ends of the distribution, which was an issue with the female sample.

Figure 2.6: Change in Interquartile Range by Gender



(a) (b)

Figure Notes: The change in interquartile covers the change from 1984 to 2014 using the Median Restriction. The limits of the box refer to the upper and lower bounds and the whiskers are the 90% confidence interval for the change. The corresponding table with all the results of all restrictions can be found in the Results Appendix in Table A.3

and panel (b) the same but for daughters. The limits of the box refer to the upper and lower bound to the change in IQR, whilst the whiskers present the corresponding 90% confidence interval for the change.

The only statistically significant changes in the interquartile range occurred for sons of middle and high income parents. Suggesting that there has been growth in the spread of the distribution over time and increasing inequality levels for males overall. However, this does not appear to be the case for daughters. We are not able to conclude whether there has been a change in the within group inequality levels here. To further understand which group is driving changes we go on to decompose changes by education.

The Role of Education

Firstly looking at the pooled data, Figure 2.7 plots the change in the College Wage Premium over time by parent income group. We define this as the difference in the median wage between the college and non-college children. There are no obvious groups of growth, with the middle income parent group experience growth in the College Wage Premium in the early 1980's before levelling off. The most we can say is that the median wage for those with college from all parent income groups is higher than those without college.

Figure 2.7: Bounds to the College Wage Premium.

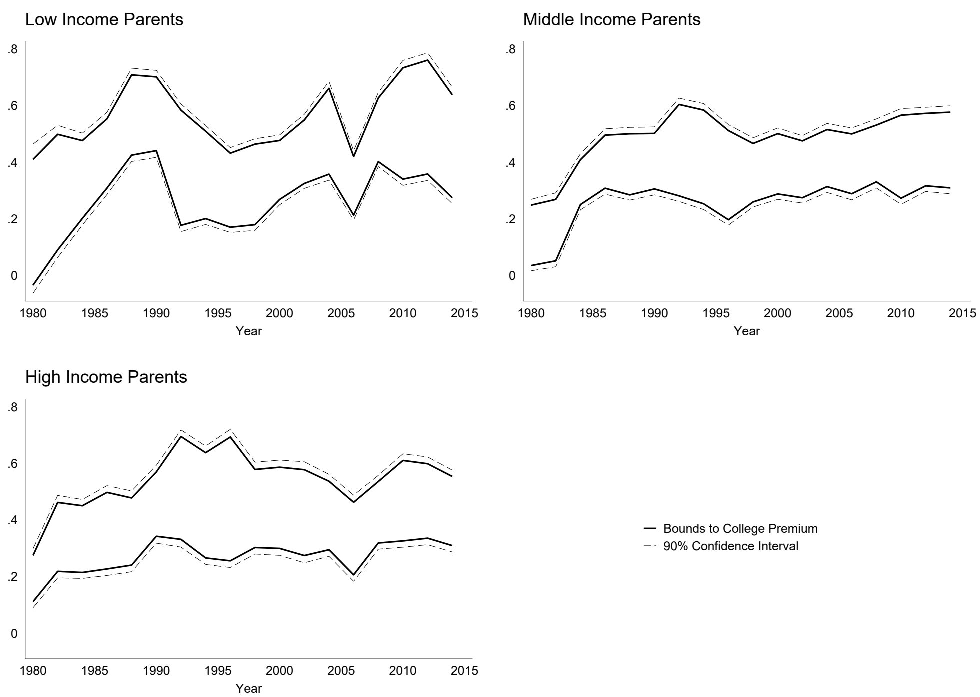


Figure Notes: On the vertical axis is the change in College Wage Rremium based on the difference in median wage. On the horizontal axis is time. Estimates are produced with the Median Restriction assumption along with the 90% Confidence intervals.

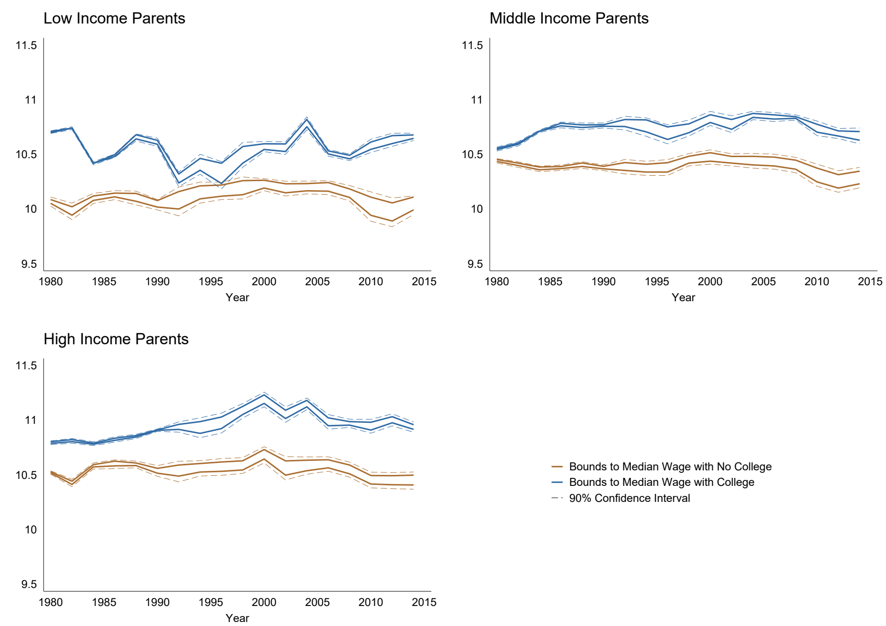
A possible cause for the strength of the growth of daughters earnings compared to sons could be their increased participation in higher education, in an attempt to close attainment gaps. We now separate out the trends for those with some college and those without by gender.

Separating out by gender, Figure 2.8 plots the median wage of both with and without college for son's in the top panel, labelled (a) and daughter's in the bottom panel, labelled (b). Again, unsurprisingly we see the dominance of having college over not having college in all cases. However, we see that for the daughter's there is growth in the median wage for both education types. This is not the case for Son's. All series for son's in panel (a) appear to have more stagnant wage growth. However, apparent substantial dominance of median wage for Son's with college suggests that the return for son's is greater than that for daughter's of investing in higher education.

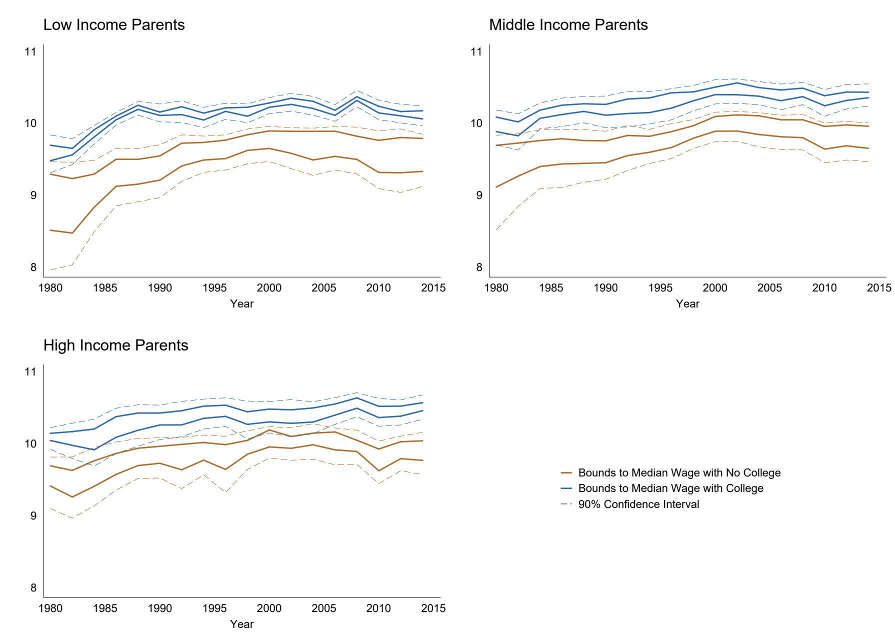
Whilst, there does not appear to be a difference by parent income group for son's in terms of college, there appears to be a substantial difference in the no college case. Son's of higher income parents appear to have a higher wage (fluctuating around 10.5) than the son's of low income parents (with values around 10) even if both groups have no college education. This story appears to be similar for daughters, however not as striking due to wider bounds in this case. This suggests there are inherent rewards to being a child of a high income family, however this reward does not appear to have been growing for the son's in the same way as for daughter's.

We now move on to focus only on the children with college, the median wage is plotted over time by parent income group in Figure 2.9. Immediately we see that the median wage of those with college has consistently been higher for son's than for daughters. Additionally, sons of high income parents seem to be the higher earners in terms of median wage. There is some evidence that this difference in median wage between sons and daughters is closing; daughters incomes seem to be converging to the son's income for all parent income groups, with the exception of the low parent income group which seems to be experiencing minimal growth in daughters median wage since 1985.

Figure 2.8: Bounds to the Median Wage by Education Level.



(a) Sons



(b) Daughters

Figure Notes: On the vertical axis is the log median annual wage by year on the horizontal axis. Estimates are produced with the median restriction assumption along with the 90% Confidence intervals.

Prior to 1986, for the low income parent group in Figure 2.9, between 1980 and 1985, daughters experienced tremendous growth in their median wage, and appeared to converge to the median wage rate of son counterparts. However, since this point there appears to have been minimal to no growth in the median wage and a possible divergence in the median wage of sons and daughters.[12] Daughters of other parent income groups also appear to have experienced growth in median wage at the start of the period studied, however this growth does not appear to be as strong or as fast as for the low parent income group.

Figure 2.9: Bounds to the Median Wage for Son's and Daughter's with College.

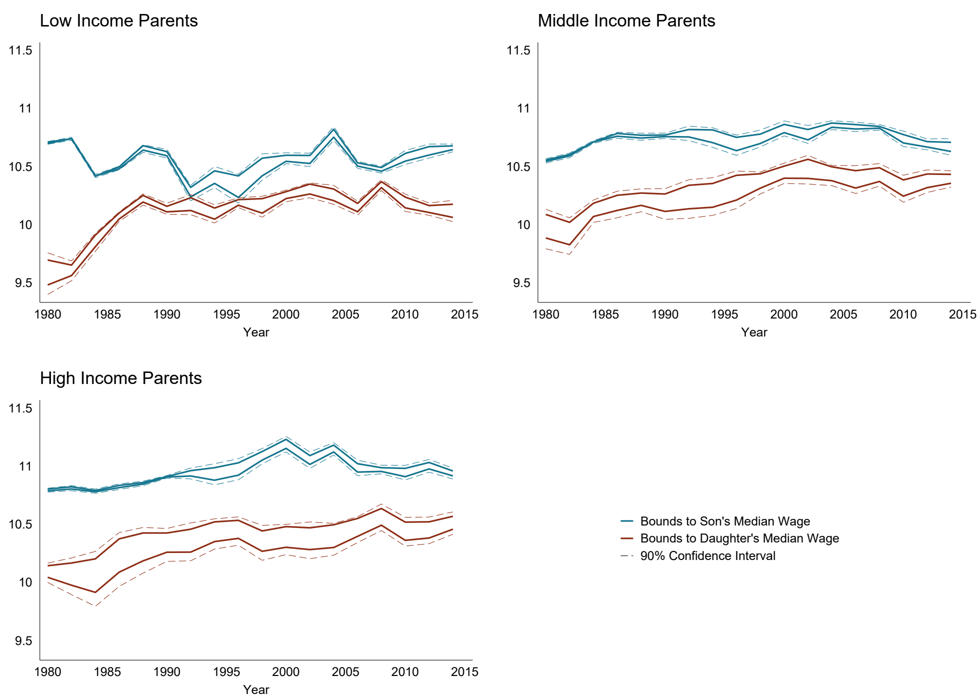


Figure Notes: On the vertical axis is the log median annual wage by year on the horizontal axis. Estimates are produced with the median restriction assumption along with the 90% Confidence intervals.

In Figure 2.10 the change in the median wage is plotted for all decompositions of gender and education for the period 1980 to 2014. Recall, that the limits of the box

[12]There is a similar story if one chooses to look solely at the children without college as presented in Figure A.2 in the Results Appendix. Here we see again that daughters median wage appears to be converging to that of the son counterparts.

refer to the upper and lower bound and the whiskers represent the 90% confidence interval for the case of the median restriction.

Firstly, we see that daughters, regardless of education level, have experienced growth in the median wage which is statistically significant for all parent income groups.[13] This is in stark contrast to the change for son's, where the no college group have experienced a decline in the median wage, along with the sons with college of low income parents. Son's of middle and high income families with college have experienced median wage growth which is statistically significant. However, this growth is not substantial in comparison to their daughter counterparts.

The group which has experienced the most growth in their median wage is the daughters with college from low income families with growth of around 0.5 log points. This is in contradiction to son's, who have no statistically significant change. This supports a hypothesis that one of the more forgotten parts of society is working class men. They do not appear to have benefited to the same extent from access to college as their female counterparts.

Overall, we can conclude that whilst education is important for all, there is evidence of differential benefits. Son's with college still earn a consistently higher wage than daughters for all parent income levels. However there is huge growth in daughters earnings with college for all parent income groups. Son's with no college, or with college from a low income family seem to be experiencing falling median wage, suggesting that the fact their wage is not growing at the same rate as the female counterparts.

[13] With the exception for the children of low and middle income parents under this conservative median restriction.

Figure 2.10: Bounds to the change in Median Wage by Education Level and Gender

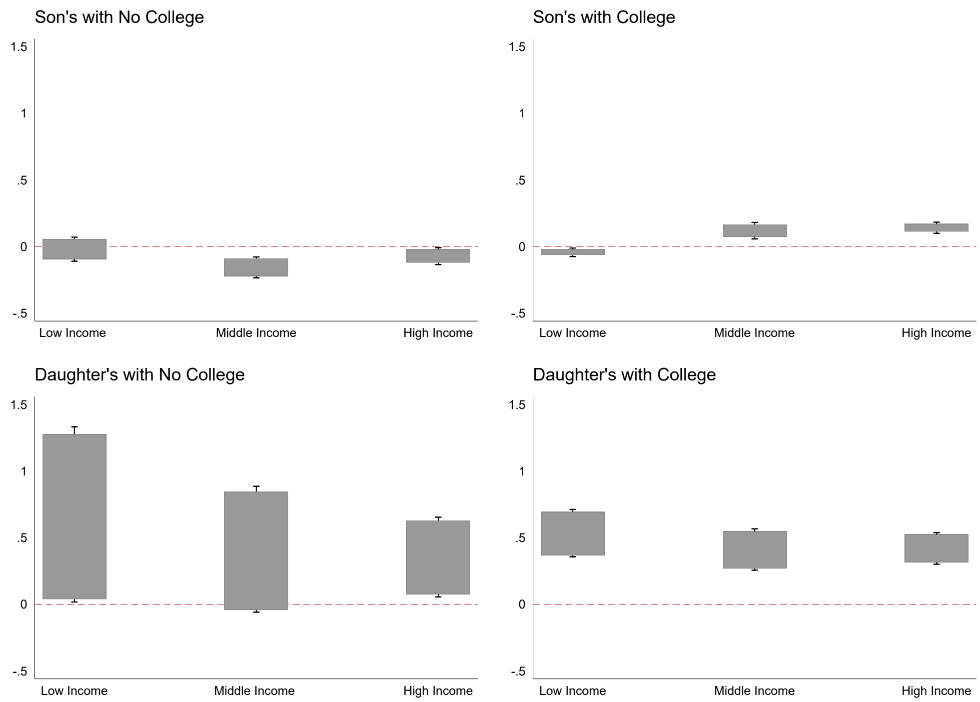


Figure Notes: On the vertical axis is the log median annual wage wage change between 1980 and 2014. The limits of the box refer the the upper and lower bounds and the whiskers refer to the 90% confidence interval. Estimates are produced with the median restriction assumption. The corresponding table with all the results of all restrictions can be found in the results appendix in Table A.4.

2.5.5 Validity and Robustness

Testing Validity

We follow the approach of Blundell et al. (2007) to test the validity of the MTS assumptions (Stochastic Dominance and Median Restriction). This involves estimating a simple mincer earners equation, and then estimating the residuals for firstly the group of workers who do not experience any time out of the labour market whilst they are part of the panel, and secondly, a group who experience some time out of the labour force whilst in the panel. For the individuals who are never in the labour force during the panel are dropped from the sample.

The residuals from the regression are then calculated for this group. The distribution of these residuals for the two groups are plotted in Figure 2.11. As is evidence by the dominance of the *always in labour market* group, the selection restrictions are valid in this case.

Figure 2.11: Distribution of Estimated Residual Wage by Labour Market Consistency

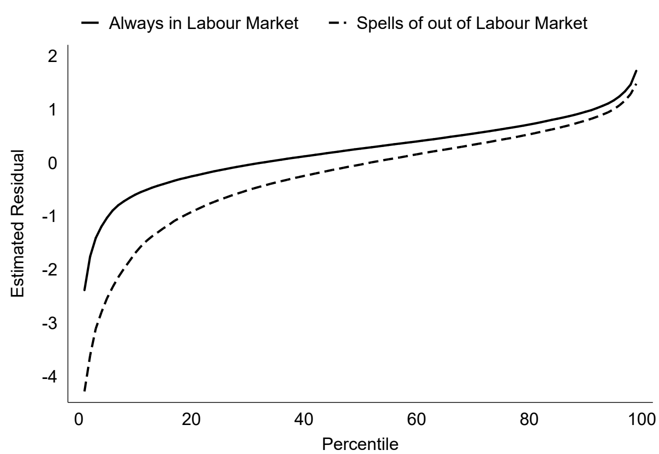


Figure Notes: Residuals for those who are always in the labour market and those who have some spells with no income at some point during the panel. Those who are never in the labour market are not included. Sample includes individuals aged 25 to 65.

Robustness Checks

We consider a number of robustness checks which involve firstly changing how we define low, middle and high income parents. Secondly, we can include the SEO sample from the PSID in our estimations (which we state as *with SEO Sample*). Recall that our results are based on the core PSID data and we define high income parents to be those with an average income in the top quartile and low income to be those in the bottom quartile. We change these in two ways: we can classify the high income parents as those in the top 35% of the distribution and the low income in the bottom 35% of the distribution (we label this as *35-65 Split*). Alternatively we become more strict in our definition of low and high income parents by considering the top 15% to be high income and the bottom 15% to be low income (we label this as *15-85 Split*). Of course, middle income in all of these case are parents between the cut-off's.

We focus on presenting robustness results under the median restriction as this assumption forms the basis of most of the results presented previously. Looking at the growth in the median wage by parent income group and our robust classifications in Table 2.4, demonstrates that regardless of the classification of the high income group they are the only group to experience growth in their median wage. The magnitude of the change is not vastly different to what our core results suggest, thus demonstrating that the choice of definitions are not important for the trends that we are reporting.

Table 2.4: Robustness Checks to the Change in Median Wage

	Low Income	Middle Income	High Income
With SEO Sample	-0.24, 0.28 [-0.26, 0.30]	-0.12, 0.22 [-0.14,0.24]	0.11, 0.29 [0.09, 0.31]
35-65 Split	-0.11, 0.31 [-0.13, 0.33]	-0.07, 0.21 [-0.09, 0.23]	0.14, 0.33 [0.12, 0.34]
15-85 Split	-0.05, 0.38 [-0.07, 0.40]	-0.07, 0.21 [-0.09, 0.23]	0.19, 0.37 [0.17, 0.39]

Note: The 90% Confidence Interval for the change is given in square brackets, all above results are using the Median Restriction Assumption.

2.6 Conclusion

The aim of this paper was to understand the social mobility patterns by gender and how these have changed over time using an agnostic estimation method, which reconciles issues of sample selection. We estimate bounds to the distribution of wages given parents income using linked parent to child data from the Panel Study of Income Dynamics. By doing so, we study the evolution of trends overall, as well as by gender and education level, conditional on parents income.

We find that whilst there has not been much change to apparent within parent income group inequality overall, there is evidence of between group changes. With the children of high income parents experiencing stronger growth in their median wage since 1980. This appears to be suggestive of an increasing divergence between the low parental income children and the high parental income children. Not only is this reflected in the changes in the median wage but also the returns to college. Children of high income parents with college can expect a greater median income than their low parent income counterparts.

When we look at the patterns further by gender we see increasing within group inequality for sons rather than for daughters. What we find is increasing wages for all daughters over the time period we consider, and by college or no college regardless of parent income group, whilst the counterpart sons appear to have more stagnant incomes. This convergence in sons' and daughters' earnings appears to have happened more quickly for the children of low income parents with the bounds on the median crossing by the end of our sample period.

Our findings are robust to different cut-offs in the classification of low, middle and high income parents, in addition to the inclusion of the SEO sample of the PSID.

Our test for validity shows that the median and stochastic dominance restrictions cannot be rejected, suggesting there is positive selection into the labour market.

Whilst our approach does not allow us to find evidence of a mechanism, evidence suggests differential benefits of a college education. Our holistic approach can lead to a believable assessment of the time trends due to us not having to comment on the selection mechanism, as would be required with more traditional parametric approaches. We implement a novel Instrumental variable in the form of mothers labour market attachment to tighten the bounds and as a result use the inference procedures outlined in Andrews and Shi (2013). A consequence of our method is that we can be sure the changes we see are not the result of changing labour force composition.

Chapter 3

Inequality in an Equal Society

The most equal society will exhibit a substantial degree of income and wealth inequality. Even in the absence of differences in talent, individuals approaching retirement will be substantially wealthier than those who are younger. Moreover, experience and seniority mean that older workers will have higher wages than their younger colleagues. Jointly, such life-cycle aspects of income and wealth give rise to a degree of inequality that is *natural* in all societies – even if each individual over the course of the life-cycle is exactly the same as any other individual.

An early version of this argument was made by Atkinson (1971), who suggested that the distribution of wealth should be expected to be unequal solely due to differences in accumulated savings over the life-cycle. In another important contribution Paglin (1975) argued that standard Lorenz curves 'combine and thus confuse' expected life-cyle inequality with other sources of inequality and proposed that the Gini coefficient should be corrected for the age structure inherent in income and wealth profiles.[1]

[1] Paglin's measure was subsequently refined by Atack and Bateman (1979) and Formby and Seaks (1980) such that it had a similar interpretation to a normal Gini coefficient, the measure we will use.

A powerful new body of evidence (particularly Piketty (2003), Piketty and Saez (2003a) and more recently, Atkinson et al. (2011), Piketty and Saez (2014) and Saez and Zucman (2016)) has transformed our understanding, and highlighted the societal implications, of long-term trends in inequality. However, following Atkinson (1971) and Paglin (1975) it is important to understand the extent to which these trends reflect changes in *natural* inequality due to changes in nations' demographics.[2] This paper addresses this need by taking the life-cycle argument to the data. The main contribution of this paper is descriptive. We assemble comparable time-series describing the long-term evolution of the Paglin-Gini for a number of countries. In doing so we document how much of the variation in income and wealth inequality over time and between countries is due solely to life-cycle effects and by implication how much reflects other factors. We then study how future demographic changes, particularly the ageing of the Baby Boom generation will impact upon inequality over coming decades.

Firstly, with detailed micro-data for the United States and then moving on to use harmonised micro-data for other developed countries (including the US), we analyse the degree to which even in the absence of any inequality between individuals of the same age group, societies exhibit substantial degrees of income and wealth inequality. In particular, restricting our sample to working age men so that we can abstract from changes in labour market participation, we show that the level due to life-cycle effects only (*natural* inequality) accounts for around one third of income inequality in the United States, with the remaining two-thirds attributable to differences between individuals, the effects of institutions, and so forth. Moreover, between the early 1970s and the early 1990s, the level of *natural* inequality increased by around 2 percentage points from just under 18 percent. The mid 1990s

[2]We refer to this component of overall inequality as *natural*, since given variation in income or wealth over the lifecycle it is unavoidable. We discuss this point further in Section 3.1.

marks a turning point, *natural* inequality declined slightly, however this has been more than offset by large increases in *excess* inequality, that is inequality attributable to other sources. This is in contrast to the other countries we study where the level of excess inequality is often lower and with a less pronounced upwards trend. Taking into account the role of demographics suggests that variations in inequality across countries are considerably larger than otherwise measured and increasing. Results for wealth show that natural wealth inequality has varied little over the last 20 years in the US as observed inequality has increased rapidly. However, life-cycle effects can explain a considerable amount of the cross-country variation in wealth inequality.

We utilise harmonised micro data from the Luxembourg Income Study (LIS) and the Luxembourg Wealth Study (LWS) for our analysis. Importantly for our purpose, these studies contain data which have harmonised variable definitions to allow meaningful comparison across countries as well as over time.

Our aim of quantifying the effect of changes in demography on inequality is similar to that of the early work of Mookherjee and Shorrocks (1982). Like them we will use the Formby and Seaks (1980) modification of the Paglin-Gini. Despite only very limited aggregated data they were nevertheless able to provide evidence that rises in inequality in Great Britain over the period 1965-1980 could be almost entirely attributed to increasing *natural* inequality. A key advantage of the much improved quality and coverage of harmonised data now available, is that we can see this trend in its proper historical context – as a temporary phenomenon soon to be reversed.

There has been relatively little recent work looking at the role of demography in inequality. Thus, by documenting the relationship between the demographic structure and the natural rate of inequality we contribute to the important recent literature on trends in inequality. We assess the impact of the disproportionate size of the Baby Boom generation on natural inequality and study how natural inequality should be expected to change, *ceteris paribus* as the demographic structure converges to its long-run equilibrium. This exercise suggests that the bulge on the demographic pyramid generated by the Baby Boom is depressing natural inequality. Hence, in the future, as the demographic pyramid settles into its long-run equilibrium, wealth and income inequality will increase. Perhaps worryingly, this process will accelerate further the trend of increasing inequality documented by the seminal contributions of Piketty (2003), Piketty and Saez (2003a), Atkinson et al. (2011), Piketty and Saez (2014), Saez and Zucman (2016). In that sense, our paper contributes to the extant literature on inequality trends by highlighting that demographic forces will exacerbate the upward trends in inequality this literature has identified.[3]

This paper relates to an important literature following Mookherjee and Shorrocks (1982) that focusses on how to attribute inequality to multiple sources. This is a complication we avoid given our focus only on life-cycle effects and on the Gini coefficient. A notable feature of all of this work, particularly that of Lerman and Yitzhaki (1985), Lambert and Aronson (1993), Cowell and Jenkins (1995), Bourguignon et al. (2008), is that they largely conclude that demographic factors are relatively unimportant.[4] Yet we argue, that *à la* Piketty and Saez (2003a, 2014)

[3]Other recent country-specific studies on trends in inequality include: Australia and New Zealand Creedy et al. (2017), Germany Fuchs-Schündeln et al. (2010), Italy Jappelli and Pistaferri (2010) and Sweden Bengtsson et al. (2017).

[4]Lerman and Yitzhaki (1985) introduce a method for decomposing the the Gini by income source and use it to show, for U.S. data for 1981, the relative importance of the earnings of the head of household versus that of their spouse or property income and transfers. Lambert and Aronson (1993) clarified the meaning of the residual term, identifying it as the extent to which there was a cross-over in incomes across age-groups due to within age-group variations in earnings. Cowell and Jenkins (1995) provide a method for computing the share of inequality that may be explained by within-group variation for the generalized-entropy class of inequality measures. Analysing one wave of the PSID they conclude that 'not much' of inequality can be explained by race, age, and gender. Bourguignon et al. (2008) develop a method by which differences in the distribution of household incomes across countries maybe compared and apportioned to different sources.

that there is much to be gained by a considering variation over time and this paper demonstrates that there have been substantial differences in the relative importance of life-cycle effects both over time and across countries and that these can account for a meaningful share of total inequality.

Some other recent work has sought to decompose the sources and evolution of inequality over time. Brewer and Wren-Lewis (2016) who decompose trends in UK inequality by income source and demographic characteristics to show that increases in inequality amongst those in employment have been ameliorated by relatively low unemployment, and more generous pension provision. Yamada (2012) studies the role of individual risk, macroeconomic, and demographic changes in Japan using an OLG model. Almås et al. (2011a) uses register data to study the role of the Baby Boom generation in the evolution of inequality of Norway.[5]

The paper proceeds as follows. The next section sketches the empirical argument for, and formalizes the notion of, natural inequality, and introduces the life-cycle adjusted Gini. Section 3.2 takes the notion of natural inequality to data. It focuses first on income inequality in the US, before considering a panel of countries. These results suggest, that particularly in the US, ignoring changes in natural rates of inequality over the last 20 years may mean underestimating increases in inequality. The last part of Section 3.2 shows that comparatively little of wealth inequality is due to natural inequality. Section 3.3 turns to the future and simulates the evolution of natural inequality as countries return to their demographic steady states following the Baby Boom. The results suggest that in many countries there will be substantial increases in natural inequality over the next 20 years. We close with a

Applying this method they are able to decompose the sources of differences in inequality between Brazil and the USA, showing that these are driven by greater inequality in education levels (and the returns to education), and pension incomes. Like Cowell and Jenkins (1995) they conclude that little of the difference can be explained by demographic factors.

[5] This work links to the related literature on lifetime inequality, for example Blundell and Preston (1998), Blundell and Etheridge (2010) and Corneo (2015).

brief conclusion. The Appendix summarises the data used and presents additional results.

3.1 Natural Rates of Inequality

Our focus on the level of inequality due solely to life-cycle factors is directly related to the prominent literature that studies the determinants of the distributions of earnings and wealth. For example, Huggett et al. (2011) consider how shocks received at different life stages affect lifetime income. The distribution of wealth is studied by Cagetti and De Nardi (2006) who study a quantitative model of occupational choice with the potential for entrepreneurship and study the role bequests and restrictions on investment play in determining wealth inequality. See also Neal and Rosen (2000) for a review and Huggett et al. (2006) for a more recent example attempting to match the extent to which more or less sophisticated life-cycle models can explain observed income-inequality. In this class of models life-cycle inequality is determined by the choice of parameters, often calibrated to US data, and the form of the model. As in Cagetti and De Nardi (2006), this approach allows for sophisticated analyses of the interaction of different features of an economy but any estimates depend on how well the model corresponds to reality and how precisely the parameters are chosen. Our approach is different, we use micro-data to study the empirical importance of life-cycle inequality for income and wealth without recourse to additional assumptions. One way we contribute to this literature is by providing empirical evidence as to the extent to which income and wealth inequality should be attributed to life-cycle effects in this type of model.

To fix ideas we follow Atkinson (1971) and start with a stylized exposition of the levels of income and wealth inequality that would prevail if the only difference between individuals is that they are at a different stage of their life cycle. Starting with income inequality, consider the following process of labour income:

$$W(v,t) = E(t-v)w(t), \tag{3.1}$$

where $W(v,t)$ is the income at time t of an individual born at time v, $w(t)$ is the economy wide wage rate and $E(t-v)$ is an individual scaling factor that creates a life-cycle pattern in labour income. $E(t-v)$ can be driven by many factors, which, for the sake of brevity we do not model separately. Indeed, for the current purpose it suffices to acknowledge that $E(t-v)$ can contain experience effects by which more senior workers earn more than junior workers but also institutional factors such as a social security system that redistributes income from workers to retirees.

This makes clear the argument of Atkinson (1971) and Paglin (1975) that the standard egalitarian view of complete income and wealth equality implies either substantial redistribution from old to young, or that there is no return to experience, etc. Indeed a society in which one never accumulates assets or develops is quite alien. This implies, as argued by Paglin (1975), that the correct benchmark is the level of inequality due only to life-cycle effects.[6] Thus, we refer to such innate earnings differences as the level of *natural* inequality.

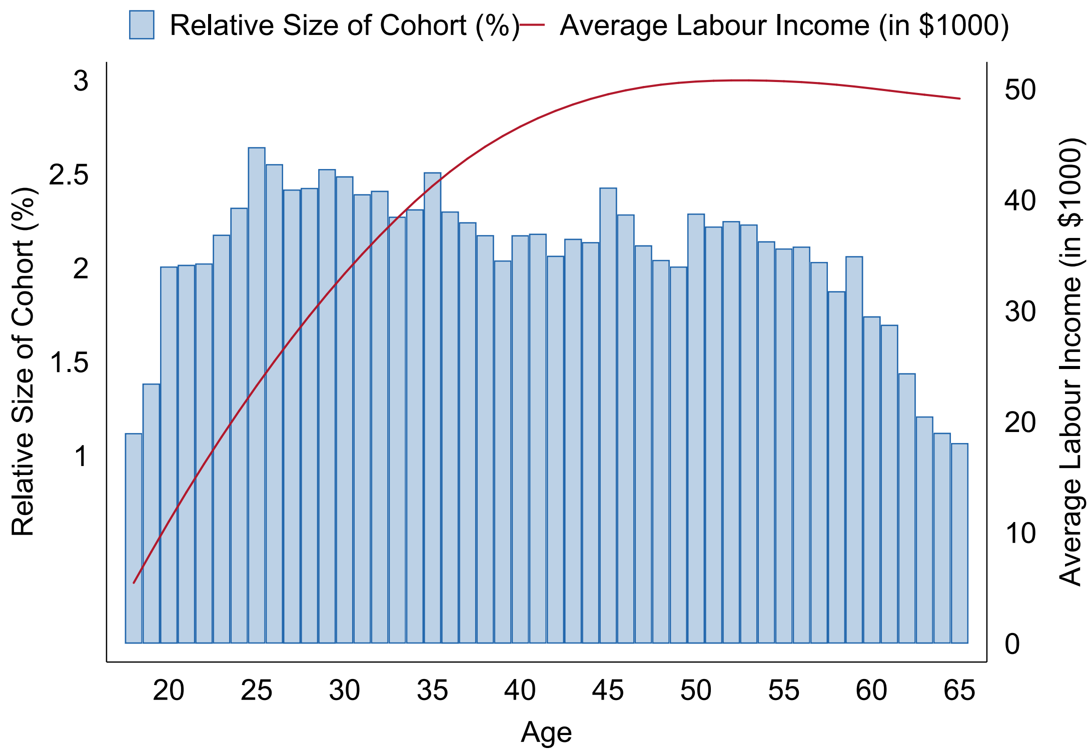
For this concept to be meaningful we require that the earnings process is relatively stable overtime, and that the age-profile of earnings is not driven by policy. It is of course, the case that there have been changes in the former, and that these partly will reflect the latter. But, as an empirical matter, as can be seen in Figures B.1 and B.2 in the Appendix, the average earnings profile has been relatively stable. Likewise, while the measured level of *natural* inequality will be affected in the short-term by cyclical fluctations in asset prices and wages. Our focus is on the long term trends, rather the smaller short-run fluctuations. [7]

However, the degree of inequality is determined not only by how much richer the old are than the young, but their relative number. The demographic structure of the UK in 1969, as analysed by Atkinson (1971), is both quite different to that of

[6]The Paglin Gini differs from other modifications of the Gini in that it maintains the same egalitarian benchmark. Other approaches include that of Almås et al. (2011a) who provide an alternative adjustment of the inequality measures, focusing on *unfair* inequality. This approach replaces the assumption incarnate in the standard Gini index or Lorenz curve that fairness implies complete egalitarianism with a more general framework that better corresponds to intuitive and philosophical conceptions of a fair society. For example, *unfair* inequality may see as fair that those who work harder or who are better qualified earn more. In their empirical analysis Almås et al. (2011a) use rich micro-data to study departures from the *fair* income distribution for Norway. Generalising standard approaches to other definitions of inequality extends in important ways our toolkit but is quite different to the approach of our paper, which maintains the standard egalitarian definition of inequality. It is also quite different in practical terms, as a key advantage of our measure is that it can be derived without having recourse to registry data with variables such as IQ, thereby enabling us to compare excess inequality internationally. We only need data on ages and income/wealth and not the detailed data used by Almås et al. (2011a). More similar to this paper is Almås et al. (2011b) who propose an alternative method of adjusting the Gini coefficient for life-cycle effects, that can better account for correlations between, say age and education levels. This is a substantial advantage, but again necessitates detailed micro-data normally not available such as parental earnings, that the effects of age and other factors may be precisely estimated.

[7]The term ‘Natural Inequality’ is used in a different, but related, sense in Philosophy as describing inequality due differences in innate characteristics rather than societal factors. For example, Gorr (1983) discusses whether Rawls’ presumption in *A Theory of Justice* that it is possible to equalise inequalities due to differences in ‘intelligence, talent, and so on’ without violating individuals’ right to express their *essential nature* is justified. From this perspective the view that inequality due to differences in age is *natural* while other differences are not can be seen, informally, as an argument that the lifecycle is essential to our *essential nature* but that differences in talent, etc., are not.

Figure 3.1: Income and Cohort Size by Age Group United States, 2015



Source: ASEC Supplement of Current Population Survey
Notes: The left y-axis corresponds to the relative size of each age cohort for men in 2016, represented by the light blue bars. The right y-axis in the average labour income in $1000 dollars for each group. Thus the red line maps the average earnings profile. The bulge in the relative population size around ages 45 to 60 is the impact of the Baby Boom generation distorting the standard demographic pyramid.

today given improvements in longevity but is also different to that elsewhere, then and now. We develop this intuition by sketching out the profile of income and cohort shares for the United States using data from the Current Population Survey (CPS). The income profile, contained in the solid line of Figure 3.1, reflects the average income of men in each age group. There we see that income has the familiar hump-shaped profile. The bars in Figure 3.1 trace out the associated cohort sizes by age. This provides the relatively uniform demographic pyramid associated with high income countries. However, in contrast to a steady-state demographic structure, where we would expect a smooth decrease in cohort size as age increases, we notice the ragged structure of the triangle - due to, for instance, the Baby Boom. Importantly, we can combine the income profile and the size of the cohorts in Figure 3.1 to calculate a Gini coefficient. This simply involves using cohort averages, $\bar{x}_i$ and $\bar{x}_j$ for each pair of cohorts i and j in place of individual data, and weighting by cohort sizes p_i and p_j, in an otherwise standard expression for the Gini coefficient:

$$\theta^{NR} = \frac{\sum_{i \neq j} p_i p_j |\bar{x}_i - \bar{x}_j|}{2\bar{\bar{x}}}. \tag{3.2}$$

where $\bar{\bar{x}}$ is the populaton mean. This provides a value of 0.16, thus attesting to the idea of a natural level of income inequality. For wealth we provide a similar analysis in Figure 3.2 where we sketch out the age profile of mean wealth for the United States using data from the Luxembourg Wealth Study. If anything, the wealth profile is more hump-shaped over the life-cycle. This translates into higher natural inequality with the Gini coefficient of wealth being 0.38.

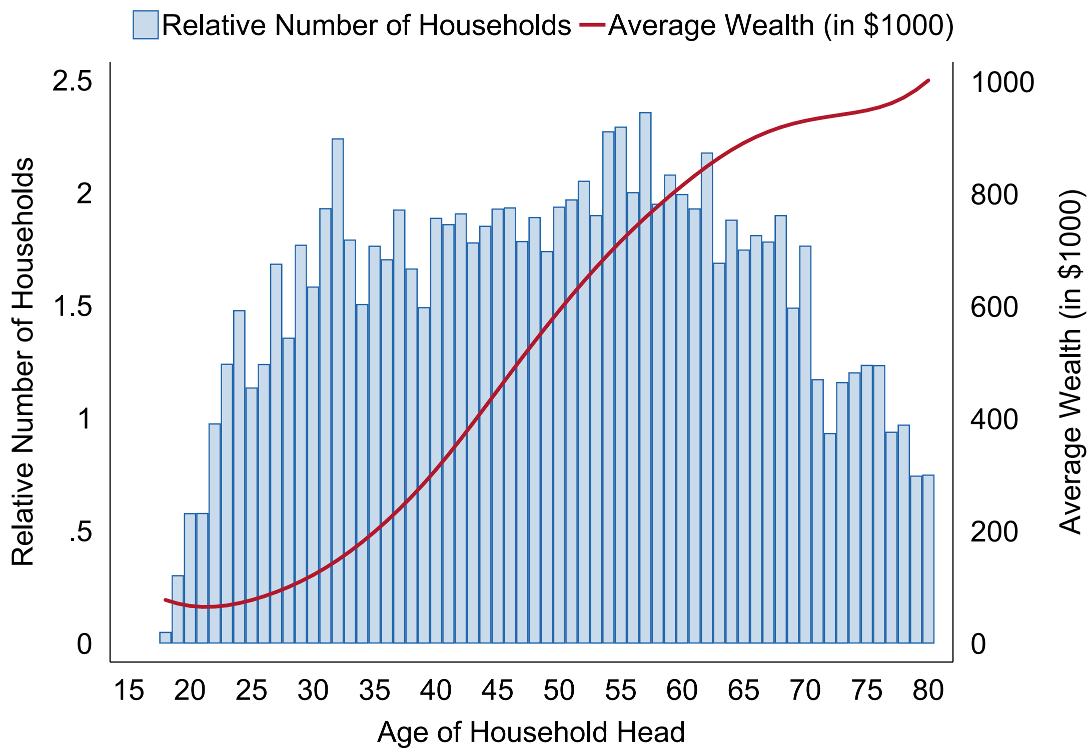
For brevity, we formalize the reasoning developed above and summarize the main conclusions from the model in the following theorem.

Theorem 1. The Gini coefficient of income (wealth) is strictly positive in the presence of a non-flat life-cycle income (wealth) profile.

Corollary 1.1. Perfect income (wealth) equality implies a flat income (wealth) profile over the life-cycle.

The proof works by writing the Gini coefficient as a product of the standardised variation of income, and the correlation of income with its rank, following Milanovic (1997), and noting that both of these terms are only zero when income is constant for all ages. The proof itself is in Appendix B.1.

Figure 3.2: Wealth and Cohort Size by Age Group United States, 2016



Source: Luxembourg Wealth Study (LWS), year 2016
Notes: The left y-axis corresponds to the relative number of households with a household head at a given age cohort, expressed by the blue bars. The right y-axis is the average wealth of each household in $1000. Hence, the red line maps the average wealth accumulation of households over the age profile of the household head. Results are produced using the household level weights.

Considering that observed inequality is generated by a host of factors, it seems appropriate to view *natural* inequality as a benchmark, deviations from which are useful as indicators of life-cycle *adjusted* measures of inequality. For expositional purposes we take a graphical approach, but there is no conceptual or quantitative difference from the approach in equation (3.2). Figure 3.3 reproduces the conventional graph defining the Gini coefficient, but with an additional Lorenz curve. The thick curved line is the life-cycle Lorenz curve – the Lorenz curve associated with the natural rate – and the dashed line is the actual Lorenz curve. A indicates the area between the line of equality and the life-cycle Lorenz curve and B and B' indicate the areas under the life-cycle and actual Lorenz curves, respectively. The natural rate Gini can be expressed as: $\theta^{NR} = 1 - 2B$, similarly the non-adjusted or conventional Gini coefficient can be expressed as: $\theta^{U} = 1 - 2B'$.[8] Using the graph we can also define the life-cycle adjusted Gini as: $\theta^{LA} = \frac{B-B'}{B}$. Which can be derived from the above Ginis as:

$$\theta^{LA} = \frac{\theta^{U} - \theta^{NR}}{1 - \theta^{NR}}. \tag{3.3}$$

implying that a society with only natural inequality will have $\theta^{LA} = 0$, while a society exhibiting inequality in excess of natural inequality will take positive adjusted values. A useful measure we will employ below is excess inequality defined as the difference between actual and natural inequality $\theta^{E} = \theta^{U} - \theta^{NR}$.

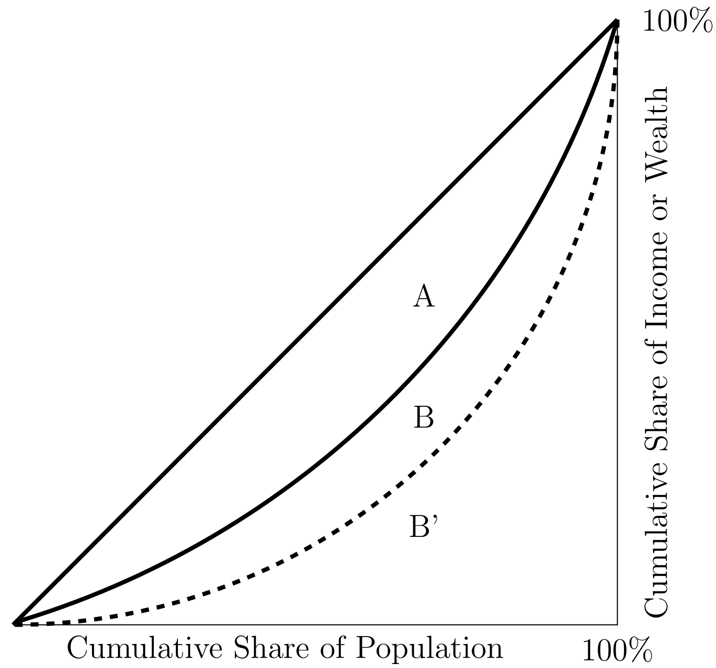
Focusing on the Paglin (1975) debate about how to properly correct for age factors in inequality, we can observe that what we call the natural rate comes closest to what he calls the A(ge)-Gini, which was not the source of controversy. In fact, it is equivalent to the Modified-Paglin Gini suggested by Formby and Seaks (1980) and also employed by Formby et al. (1989) to analyse trends in inequality.[9] We seek to build on these earlier insights by exploiting vastly improved and harmonised data to obtain precise and comparable estimates of the inequality trends of multiple countries and, importantly, to predict the development of inequality into the future.

To cement ideas, we present three measures of inequality in our analysis. Firstly, what we refer to as the *Unadjusted* Gini, which is the conventional measure of inequality which is bounded between zero and one; with higher values referring to

[8] For comparison, see the notation of (3.2), we can write the conventional Gini, computed over each pair of individuals l and k as $\theta^{U} = \frac{\sum_{l \neq k} |x_l - x_k|}{2 \sum p_l \sum_l x}$

[9] Their modification of the Paglin (1975) measure amounts to redefining the denominator of θ^{LA} as B and not $A + B$.

Figure 3.3: The Life-Cycle Adjusted Gini Coefficient



The solid diagonal line is the conventional line of perfect equality. The solid curve is the Lorenz curve associated with the natural rate. The dashed curve is the actual Lorenz curve. A is the area between the two solid lines, and B is the area under the natural rate Lorenz Curve. B' is the area under the actual Lorenz curve. The natural rate Gini can be expressed as: $\theta^{NR} = 1-2B$, similarly the non-adjusted or conventional Gini coefficient can be expressed as: $\theta^{U} = 1 - 2B'$.

greater deviations from equality. Secondly, we introduce *natural* inequality which is measured by the natural rate Gini, the interpretation mimics that of the conventional Gini; deviations from equality due to life-cycle factors. This gives rise to what we call *excess inequality* which is simply the difference between these two measures. Before, finally presenting what we call the *life-cycle adjusted* Gini which is the deviations from natural inequality rather than equality and holds the same interpretation as the conventional Gini, and is the measure initially presented in Paglin (1975) and Formby and Seaks (1980).

In taking this argument to the data one previously neglected, but important, subtlety in the computation of the Paglin Gini emerges. This is the choice of the relevant population, given both unemployment and endogenous labour market participation. If one includes the entire population as is implicit in the work of Paglin (1975) and Formby and Seaks (1980) then the income attributed to those unemployed, or not in the labour market becomes important. As is how the income from shared assets is attributed. This is true, a fortiori, for our purposes since we are making comparisons across countries and over a period in which dispersion in retirement ages has increased.

More concretely, the decision to retire embodies choices that are endogenous with respect to earning potentials as well as societal mores and institutions. For this reason, we restrict, as in Figure 3.1, our analysis to people aged 18-65 for the purposes of analysing labour income.[10] This minimises concerns about endogenous selection in to full- or part-time employment once of retirement age. As per Figure 3.2 for wealth we consider the entire population, but to avoid having to split jointly held assets, choose households as the unit of analysis.[11]

To address concerns about endogenous labour market participation at other ages our analysis will focus on natural inequality between men with positive earnings.[12] Thus, at all ages we are comparing only those in work (including the self-employed). While, it might be reasonable to presume that those who do not have positive earnings are mostly unemployed, attributing to them earnings of zero leads to estimates of income inequality substantially higher than conventional

[10]Our findings are not sensitive to this choice of cutoff. Figure 3.6 discussed below shows that including older people does not impact the key qualitative and quantitative conclusions.

[11]A related issue is how to define age-groups. In results available upon request we document that the bias of the Gini coefficient is decreasing in the number of groups, and neglible if we work with individual years. The large sample surveyed by the CPS means that sample size concerns that might have motivated pooling into coarser age groups in previous work can be ignored.

[12]While, Men retire at different ages, and average retirement ages have varied, our results are robust to a range of alternative cutoffs.

estimates. More importantly, given the purpose of this paper is to understand the relative importance of natural inequality over time, including those with zero earnings will also introduce into the calculation of natural inequality a component that is not *natural*. For example, if youth unemployment is high then including the unemployed will overstate the natural rate of unemployment by conflating the lower human capital of younger workers with the effects of other factors that are driving unemployment. Whilst potentially difficult policy challenges, such factors are not inescapable in the same way as the accumulation of skills and experience over the life-cycle is. The issue is more complicated for women as an assumption that zero earnings reflects unemployment is patently untrue. Changes in female labour market participation rates have been the largest change in the labour market over the period we study but still vary markedly across developed countries, and are changing within them, limiting what may be reasonably inferred. By focusing on the subpopulation of prime aged men we are able to abstract from this and the other key labour market changes of the period. These were the increase in the share of University Graduates and Skill-biased Technological Change. We include students in our sample, as to exclude them would potentially bias our estimates as it would increase the average income of the young since they are more likely to be students. Thus, changes in student numbers might alter the average life-cycle income depressing average incomes in the first few years of adulthood and raising them in later years. We note however, that there do not seem to be substantial changes in the life-cycle earnings profile over the period.

There is of course a trade-off incarnate in restricting the sample we consider. By excluding the elderly we restrict our attention to total and natural inequality amongst those of working age, ignoring the important consequences for total inequality of longer lifespans and changes in pension provision. By excluding women we exclude the important impact that women's increased participation and equality in the labour market will have had. We argue that this is the necessary cost of ruling out the effects of endogenous responses to other changes in society. As well as highlighting the challenges in taking a longitudinal approach, we argue that this also highlights the importance of not relying on a cross-sectional snapshot to infer the relative importance of demographic characteristics in explaining inequality.

Another important aspect to note is that we cannot say anything regarding the changing composition of the work force in terms of their industry or heterogeneity within age group in terms of type of work. Inequality arising here would be deemed *unnatural* but may be confounded with our measure. Moreover, cross country differences in inequality which we observe might be explained by the relative size of workers in each sector varying by country and over time. We could address such concerns by producing a life-cycle adjusted Gini which defines a cohort by age and industry. This however is beyond the scope of this paper and we leave it for future work.

In sum, taking inspiration from Atkinson (1971), Paglin (1975) and Formby and Seaks (1980) this section has sought to reinvigorate the argument that a stylized economy populated by individuals who are equal to each other at every stage of the life-cycle displays a substantial degree of income and wealth inequality. Moreover, we have seen that this measure can be used to calculate a life-cycle adjusted Gini coefficient.

3.2 Inequality in an Equal Society

This section empirically assesses the quantitative importance of *natural* inequality. First for the United States and then for a cross-section of developed countries.

3.2.1 Inequality in the United States

For clarity, and in line with much of the focus of the literature, e.g. Piketty and Saez (2003a), Saez and Zucman (2016), we begin our analysis by focusing on the United States, using the Current Population Survey (CPS), the details of which may be found in Appendix B.1.1. We use these data in preference to the World Income Database (Alvaredo et al., 2016) because they contain the necessary detailed microdata. Similarly, using register data such as that used by Almås et al. (2011a) is infeasible because we wish to study a range of countries for a sufficiently long period. Moreover we use the CPS in favour of the LIS to maintain comparability with other the majority of other recent studies, such as Heathcote et al. (2010). The results are similar if instead we use the harmonized data of the LIS, as we will in our comparison of trends across countries in Section 3.2.2 below.[13]

The definitions of income which we use throughout are similar in both datasets. For the CPS, labour income is the total pre-tax income from employment. Similarly, the corresponding variable from LIS is defined as any monetary payments received from employment. Total income is the total pre-tax personal income or losses from all sources for the CPS and in LIS is described as income from labour and transfers.[14] An important note is that the CPS data are top-coded and this might lead us to understate inequality. Furthermore, the rules around the imposed top-coding procedures have changed over its sampling period. In the results presented we do not include a correction for top-coding but we obtain the same results if we instead apply the Pareto-interpolation correction suggested by Heathcote et al. (2010). We now go on to discuss the results for the United States.

[13]We present in the same results for the United States in Appendix B.2, where Figure B.3 for total income and Figure B.4 for labour income.

[14]A more complete description of all the data used is given in Appendix B.1.1.

Consider first the solid red line in Figure 3.4. This shows the Gini coefficient of labour income for the period 1961 to 2015 while the blue dashed line shows the Gini coefficient of total income for the same period. The most striking feature is the pronounced and consistent upwards trend over the period. The Gini was 0.36 for labour income and just above 0.40 for total income in 1961 and 0.47 and 0.50 respectively in 2015. Also clear, is that inequality in labour income has increased more than that of total income, with total income experiencing a less steep upward trend. For both series, it is apparent that the biggest growth in inequality was experienced in the period 1974 to 1995. While the trend is clear, there is also a substantial cyclical component, as was shown more generally by Milanovic (2016). Finally, we can note that the growth in inequality is faster from 2000 onwards for both series.

We now analyse the extent to which these changes in inequality reflect demographic changes. Figure 3.5 plots, for labour income, both actual (green circles) and natural inequality (blue diamonds), as well as our two measures of the difference: excess (red squares) and adjusted (purple triangles). As outlined in Section 3.1, the natural inequality (from which excess and adjusted inequality are derived) is calculated by determining the Gini coefficient of average incomes by age. We can see that natural inequality increased from 1961 to the late 1980's by around 8 percentage points. Before falling slightly, by almost 3 percentage points over the rest of the period to 2015.

Considering actual, natural, excess, and adjusted Ginis in Figure 3.5 together it is clear that while inequality increased only modestly from 1960 to 1990, this was in spite of a substantial increase in natural inequality. Indeed over the period 1960-1980 excess inequality declined, by the late 1970s half on inequality was natural. On the other hand, the substantial increase in labour income inequality since the

Figure 3.4: Actual Gini Coefficients for Labour and Total Income

Source: Authors' calculations using ASEC Supplement of the Current Population Survey years 1962-2016
Notes: The graph shows trends over time in unadjusted Gini. Labour Income (solid line) includes those aged 18-65 and total income (dashed line) includes those aged 18-78. For both time series we exclude individuals with a zero or negative income. Results are calculated using individual weights.

mid-1990s has been despite no increase in natural inequality. Excess inequality has rapidly increased. The difference between these two periods is important as it makes plain the quantitative importance of our argument. Ignoring the role of demographic change in generating variations in the natural rate of inequality can lead us to understate the increase in inequality over the last 25 years. Equally, it leads us to overstate it for the previous 25, and thus also to understate the difference between the two periods.

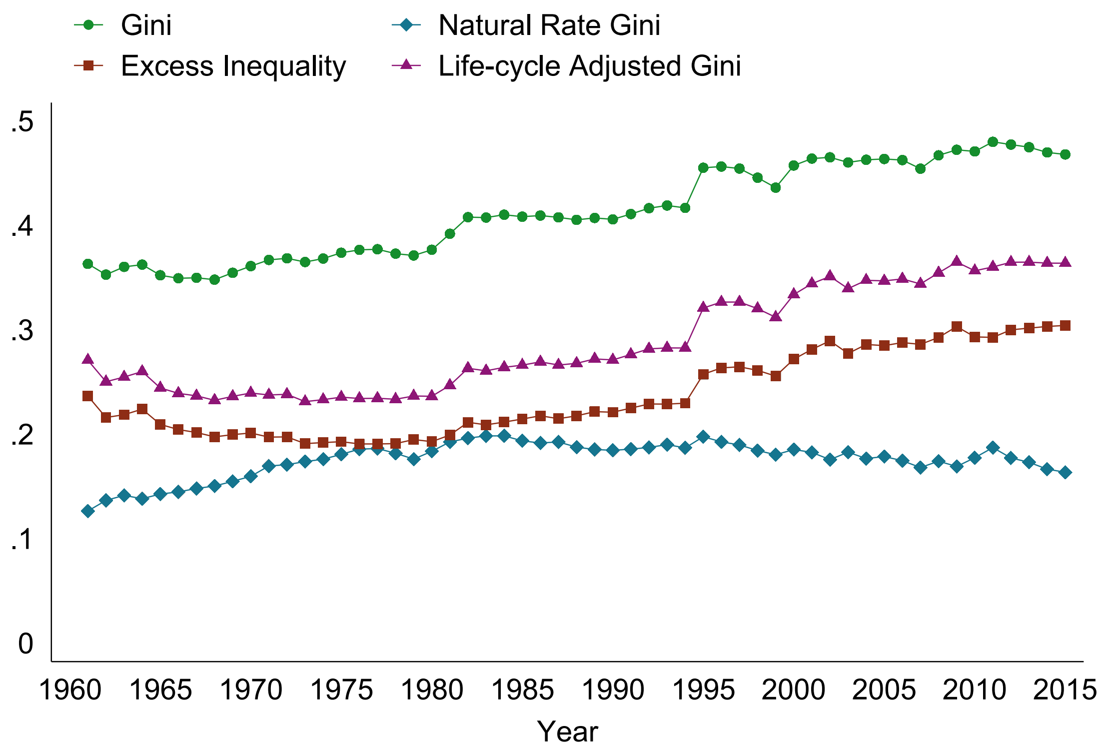
Comparison with Figure 3.6 that these results are robust to alternatively considering inequality in total income (calculated over the male population aged 18-78). In both cases excess inequality accounts for around three quarters of prevailing inequality in the US – the adjusted Gini is around 0.35 for labour income and 0.40 for total income. Moreover, trends in the two have been similar over the period with a substantial increase since the 1960s, particularly in the period since 1990. One interesting feature of the data is that the frequency with which natural and excess inequality vary are noticeably different. Changes in natural inequality are of lower frequency than changes in excess inequality which is known to be cyclical (Milanovic, 2016), perhaps as expected given the gradual nature of demographic change. Thus, changes in the natural rate are of most importance when analysing the evolution of inequality over substantial periods of time.

3.2.2 Cross Sectional Time Series Analysis

We now broaden the discussion to a sample of countries with sufficient time series available from LIS to conduct a meaningful study of trends over time. Figure 3.7 summarises the cross country variation in wave IX of the LIS for all of the countries we consider.

Natural inequality is blue, and excess inequality is red. The sum of these gives actual inequality in labour income, reported to the right of each bar. The most obvious feature of the data is the substantial variation in actual inequality, between 0.49 for the US or Canada and 0.30 for Hungary or Italy. This variation is continuous, meaning that there are no obvious 'groups' in the data. Secondly, we note that there is similarly large variation in excess inequality. For example, actual inequality in Spain or Germany is similar, but excess inequality is much higher in Spain. Alternatively, if Spain had the same demographics as the US, it would be nearly as

Figure 3.5: Actual, Natural, and Excess Gini Coefficients of Labour Income for the US 1961-2015

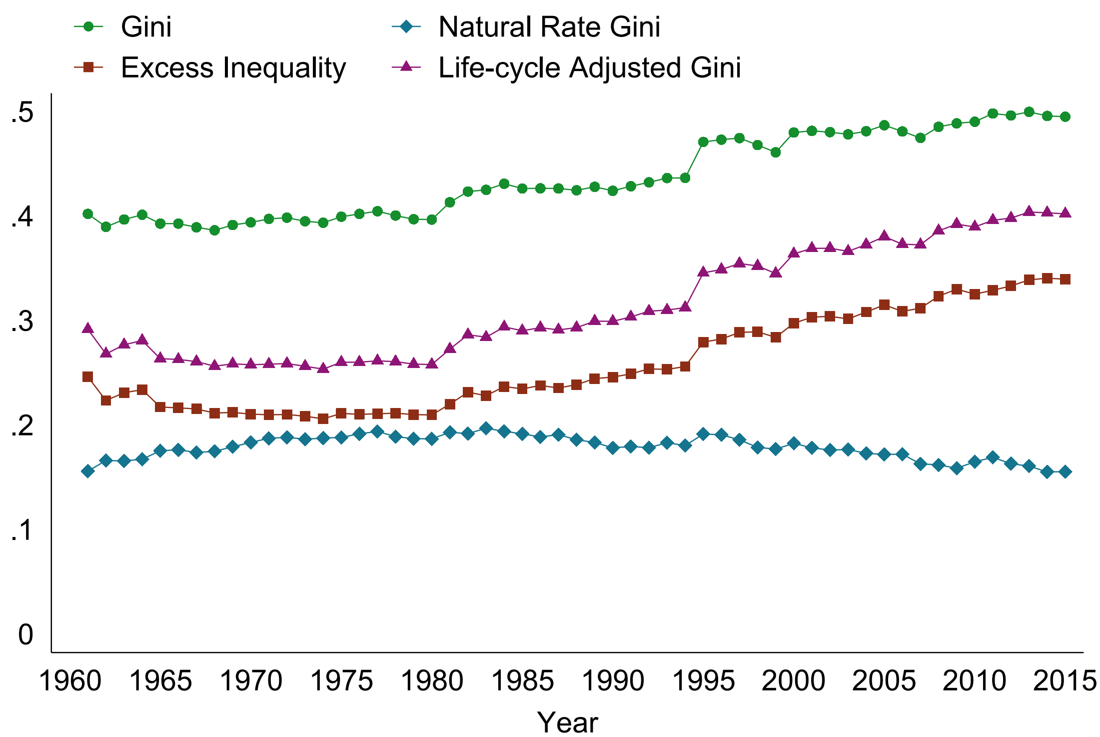


Source: Authors' calculations using ASEC Supplement of of the Current Population Survey.
Notes: Sample includes Men with positive income and are aged 18-65. Results are calculated using individual weights.

unequal. Conversely, while natural inequality in Slovenia is similar to that in Spain, excess inequality is around 7 percentage points lower. Thus, cross-country comparisons of actual inequality may be misleading. France and Finland have the same actual Gini, but excess inequality in France is higher, and thus perhaps more amenable to policy. This emphasises that as well as being important in understanding variation over time, separating natural and excess inequality is crucial to a nuanced understanding of cross-country variation in income inequality.

In moving on to consider both cross sectional and time series variation we, initially, restrict our attention to a subset of the countries for which sufficient data are avail-

Figure 3.6: Actual, Natural, and Excess Gini Coefficients of Total Income for the US 1961-2015



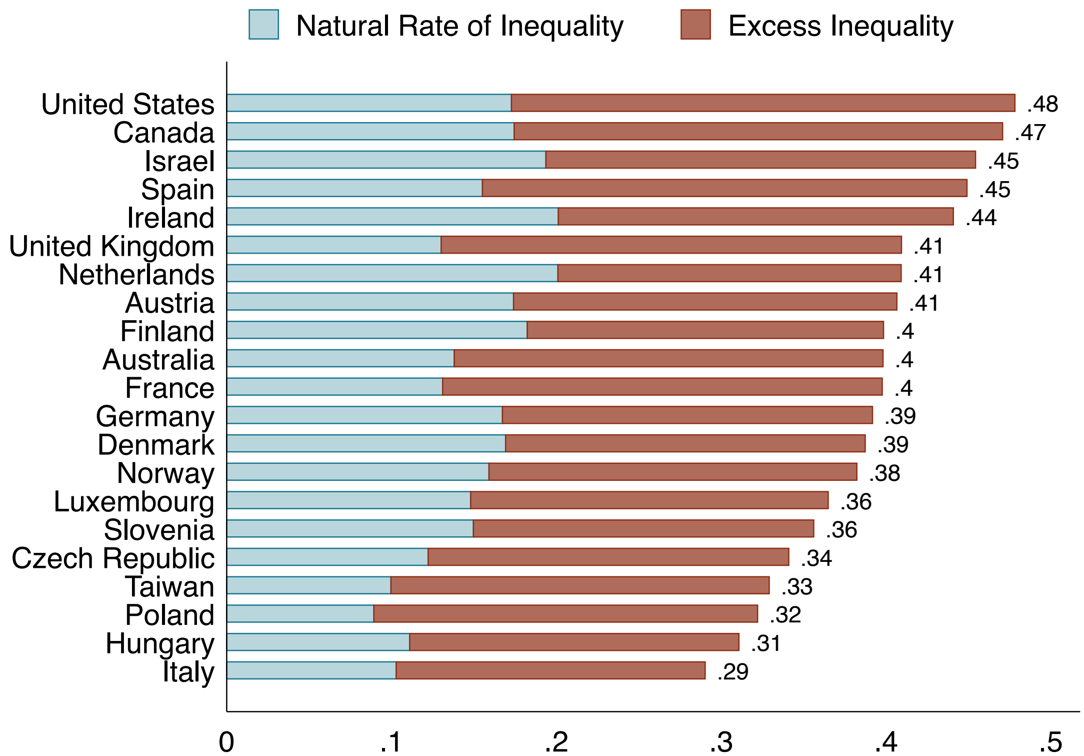
Source: Authors' calculations using ASEC Supplement of of the Current Population Survey.
Notes: Sample includes Men aged 18-78. We exclude individuals with a zero or negative income. Results are calculated using individual weights.

able in the LIS, as reported in Figure 3.7.[15] As well as focusing on those for which the data provide for a sufficient time series to look at the trends in inequality, we also limit our sample to a group of countries designed to be representative while ensuring clarity. To ensure comparability we prioritise countries for which gross income information is available. The countries which we discuss here are Canada, (West) Germany, Netherlands, Taiwan, United Kingdom and Spain.[16] The United

[15]Data are for wave IX of the LIS data, with the exception of France and Ireland where the data is for wave IIX. Mexico is excluded as the last wave available is wave VI.

[16] Results for Germany are for West Germany only throughout. Figures for Spain are for net incomes. Results for all other countries are for gross incomes. See Appendix B.1.1 for more information.

Figure 3.7: Cross Country Variation in Natural and Excess Inequality

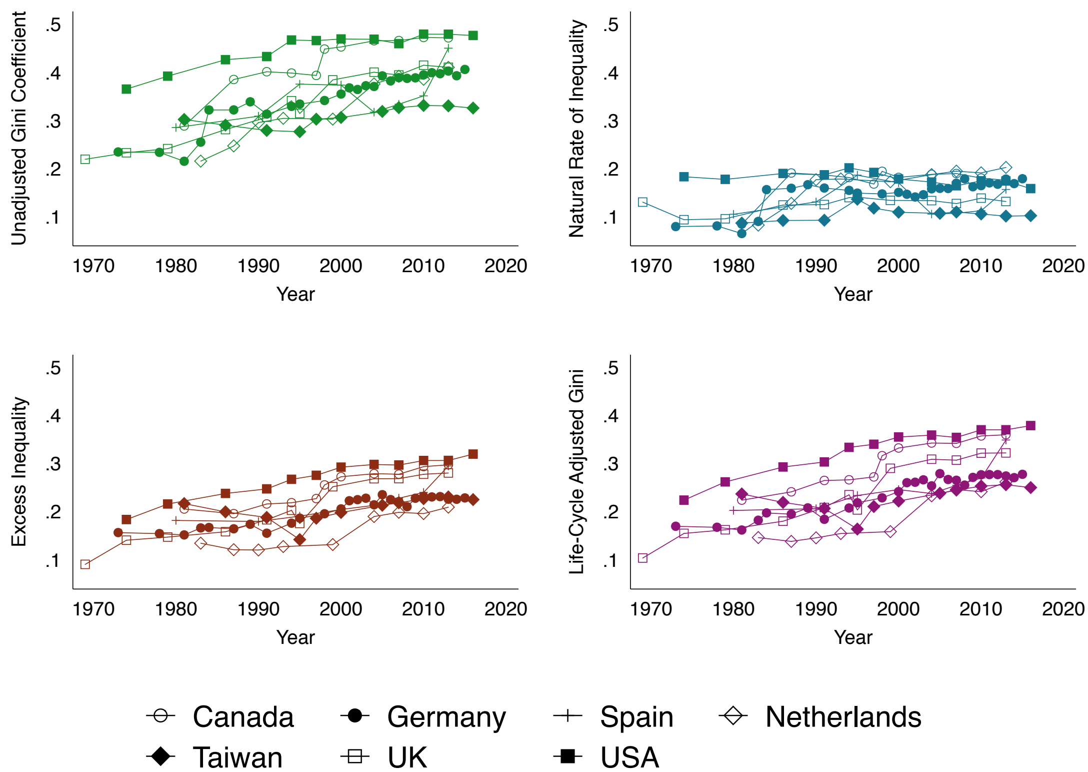


Source: Authors' calculations using LIS Wave IX, (circa 2013)
Notes: The number to the right of the bars for each country denotes the *actual* Gini, and the total length of the bar. Thus this graph shows the decomposition of the level of *actual* inequality into its *natural* component (Blue) and *excess* inequality (red). All data are for gross incomes, apart from for Israel and Slovenia which are net, and Italy and France which are mixed. Individual level weights are used in all cases. Sample includes men ages 18-65 with positive labour incomes.

States is presented again in order to make a comparison with other countries. We discuss regression analyses of the trends for the full set of countries below. Figures describing the other countries are available in the appendix.

We begin by considering labour income. Looking at the top left (green) panel of Figure 3.8, we can see that the actual Gini coefficient in the US is high compared to the other countries we consider, particularly at the beginning of our sample period. However, the gap has narrowed and all countries have experienced rising

Figure 3.8: Adjusted and Unadjusted Gini of Labour Income: Selected Countries: 1969-2016



Source: Authors' calculations using LIS data.
Notes: All results are calculated using data on gross incomes with the exception of Spain which are net incomes (with exception of wave IX). We consider those aged between 18-65 and who have positive earnings. Results are calculated using individual level weights.

inequality. Looking closer, it is clear that the biggest changes have been in Spain, the Netherlands, and Germany. In comparison, the US and Taiwan seem to have experienced relatively stable levels of inequality in labour income.

This finding is cast in new light when we consider the natural rates of inequality presented in the top-right (blue) panel of Figure 3.8. While natural inequality is stable on average, this masks comparatively notable increases for Spain, Germany and the Netherlands. This suggests that the similar trends in inequality have different sources in the US than elsewhere.

This difference is clearer when we consider adjusted inequality, displayed in Figure 3.8 in the bottom-right (purple) panel. Now we can see that the US has seen a substantial increase in adjusted inequality, both starting and finishing the period at a higher level of adjusted inequality than elsewhere. Taiwan is notable in that adjusted inequality has remained relatively stable over the sample period. Other countries, such as the the UK and Canada, have seen rapid growth rates of adjusted inequality similar to those in the US, albeit from lower initial levels. In general, the rate of increase was relatively slow everywhere until the mid 1980s after which it accelerated. The similarities in these trends, allowing for different starting points, suggests that rises in excess inequality may be driven by technological and policy changes common across the developed nations.

To demonstrate that our results are not specific to the countries plotted, Table 3.1 reports the results of estimating a linear trend using a simple fixed-effects model.[17] We report results for both total income and labour income in the first and second rows respectively. Hence, the first column reports results for the actual Gini in a model in which the trends are assumed to be homogenous across countries: $y_{it} = \tau \times t + \mu_i + \epsilon_{it}$. For both income and labour income the slope is positive and precisely

[17] Given the small number of observations, these simple estimators are preferred to more sophisticated alternatives.

estimated, reflecting the secular upwards trend in inequality. The second column reports estimates from the mean-group estimator of Pesaran and Smith (1995) in which the reported coefficients are the averages of the coefficients from separate regressions for each country: $y_{it} = \tau_i \times t + \mu_i + \epsilon_{it}$. The results are qualitatively unchanged. Inspection of the individual slopes makes clear that virtually all countries exhibit positive and significant trends.[18] This provides broader support for the previous finding of consistent upwards trends. However, as above, there are differences between labour and total income. Using both estimators, the results using *adjusted* inequality as the dependent variable suggest that, for total income, it is increasing at the same rate as actual inequality. This again highlights that the increasing importance of adjusted inequality in the US is an outlier. However, for labour income it is clear that adjusted inequality cannot explain all of the increase in actual inequality. There is a gap of between 5 (FE estimates) and 7 percentage points (MG), which suggests that around a quarter of increases in inequality have been due to demographic change.

3.2.3 Wealth Inequality

As well as increases in income inequality, the prior literature has shown that increases in wealth inequality have tended to be even larger than those in income inequality. To understand the role of demographics in this pattern, we repeat our prior analysis for wealth using the Luxembourg Wealth Study (LWS).[19] These data, like the LIS, are harmonised cross country data. Although the LWS does not have the coverage of the LIS we are able to construct a limited time series for the United States and make cross-sectional comparisons for a number of other countries, which

[18] These are reported in Table B.1 in the Appendix.

[19] Luxembourg Wealth Study (LWS) Database, http://www.lisdatacenter.org (multiple countries; 1995-2016). Luxembourg: LIS. Refer to Appendix B.1.1 for a data description.

Table 3.1: Time Trends in Inequality

	Actual		*Adjusted*	
	(1)	(2)	(3)	(4)
Labour Income	0.37***	0.39***	0.32***	0.32***
	(0.04)	(0.04)	(0.03)	(0.05)
Total Income	0.32***	0.34***	0.33***	0.32***
	(0.03)	(0.05)	(0.03)	(0.05)
Estimator	FE	MG	FE	MG
Countries	22	22	22	22
N	216	216	216	216

FE Estimator denotes the standard fixed-effects estimator with an homogenous time trend, with robust standard errors in parenthesis. MG denotes the mean-group estimator of Pesaran and Smith (1995) using the outlier-robust mean of coefficients, with standard errors in parenthesis. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

we have discussed with respect to income inequality and are available in the LWS data. The choice of data is a delicate one, the LWS data are topcoded, unfortunately the WID data (Alvaredo et al., 2016) which contain much better information on the very wealthy do not contain sufficient age data.

We choose disposable net worth (non-financial assets plus financial assets (excluding pensions) minus total liabilities) as our measure of wealth but this choice is not important for our results.[20] Wealth data are measured at the household rather than at the individual level, because of this we use the head of the household's age as a proxy, in favour of attempting to divide assets within the household. Again, this assumption does not matter for our results.

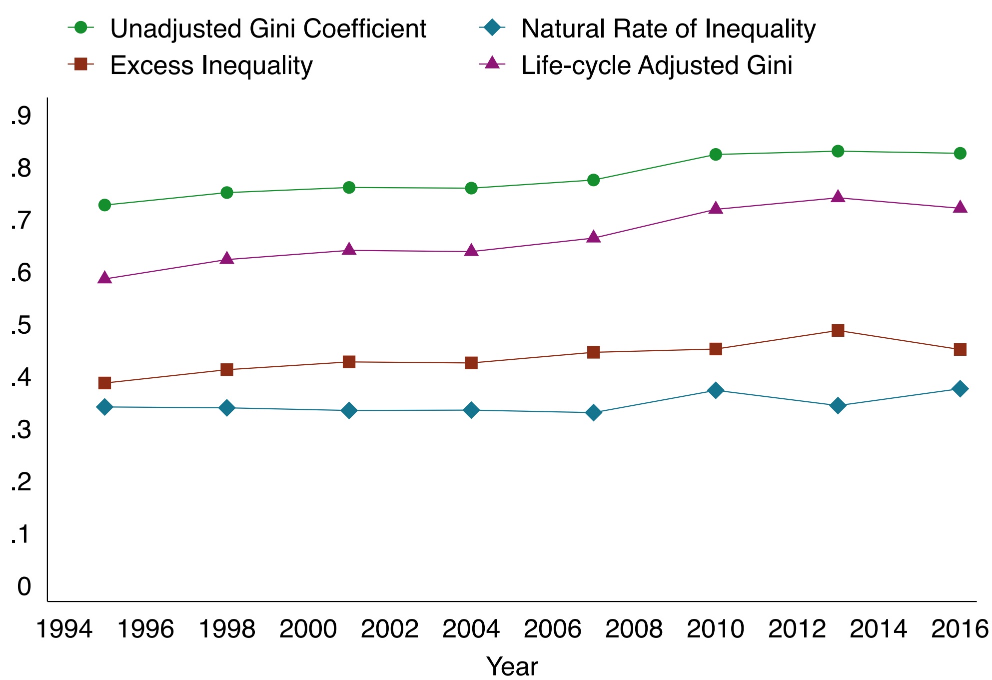
Figure 3.9 shows the (actual) Gini coefficient of wealth inequality for the United States over the period $1995 - 2016$. As expected wealth inequality is higher than

[20]We drop the top 1% of the distribution to limit the effects of topcoding procedures in the original datasets. Similar results are obtained with the alternative of interpolating the true values of the topcoded observations assuming a Pareto distribution as in Heathcote et al. (2010). This measure is preferred over others, as pension data is not as comparable across countries and for some it's not available.

income inequality over the same period. We can see that while inequality has been increasing, that changes in the natural Gini have contributed to this, although excess and life-cycle adjusted Gini have also increased. More precisely, the excess Gini of wealth has increased by around ten percentage points over the 20 year period, while natural inequality increased by four percentage points. Of course, our focus on the Gini coefficient is in contrast to much of the literature which uses concentration indices such as the share of the top 1% or 0.1%. We would not expect demographics to affect these concentration indices, but our approach here will capture changes amongst the moderately wealthy. It is clear, that whilst there is a substantial increase in the adjusted Gini that increases in natural wealth inequality have also played an important role.

Table 3.2 shows results for the ten countries for which wealth data are available. We can see that the wealth inequality varies substantially, between 0.53 in Slovenia and 0.82 in the US. However, the second and third columns suggest that this variation is in part driven by variations in the natural rate. This is 0.38 in the US but only 0.14 in Slovenia, and excess inequality is relatively consistent compared to actual inequality varying between 0.35 in Australia for the US to 0.45 in the US. Comparing the US and Canada is instructive as while the actual Gini coefficients are quite different (0.82 and 0.68 respectively) the excess Ginis are very similar (0.45 and 0.44). Thus, abstracting from life-cycle effects both societies (at least on this basis) are similarly unequal, and the US appears less of an outlier. Thus, natural inequality is arguably as or more important in understanding the cross-sectional variation in wealth inequality than it is for the time-series variation. This highlights, again, that considering the actual Gini alone may be misleading.

Figure 3.9: Wealth Inequality over Time (United States)



Source: Authors' calculations using Luxembourg Wealth Study (LWS).
Notes: Time series for United States, the underlying data are from the Survey of Consumer Finances and the wealth measure used is disposable net worth. The sample includes all households who have a head who is aged 18-78 including those who are recorded as having zero or negative net worth. Household level weights are used to produce results.

3.3 Inequality and the Baby Boom

We have seen that individual life-cycles have a central role in understanding inequality. An implication of this is that demographic dynamics will lead to changes in the distributions of income and wealth. Economists have paid considerable attention recently to long-run trends in inequality, prominent studies include Piketty (2003), Piketty and Saez (2003a), Piketty (2011), Piketty and Saez (2014) and Roine and Waldenström (2015). In this section we ask: what is going to happen to natural rates of inequality, over the next forty years as the Baby Boom generation passes, and the demographic structure returns towards its long-run equilibrium?

Table 3.2: Wealth Inequality

	Actual	Natural	Excess	Adjusted
Austria	0.66	0.22	0.43	0.56
Australia	0.58	0.23	0.35	0.45
Canada	0.68	0.24	0.44	0.58
Germany	0.76	0.27	0.49	0.67
Finland	0.62	0.24	0.38	0.50
Italy	0.55	0.16	0.39	0.47
Norway	0.76	0.37	0.39	0.61
Slovenia	0.53	0.14	0.40	0.46
UK	0.58	0.23	0.35	0.45
US	0.82	0.38	0.45	0.72

Actual is the conventional Gini coefficient. Natural, Excess, and Adjusted are the alternative measures of inequality defined in Section 3.1. Results are rounded to two decimal points. Results for Austria and Australia refer to 2014, Canada and Germany refer to 2012, Italy and Slovenia refer to 2014, Finland and Norway refer to 2013, the US refer to 2016, and the UK to 2011.

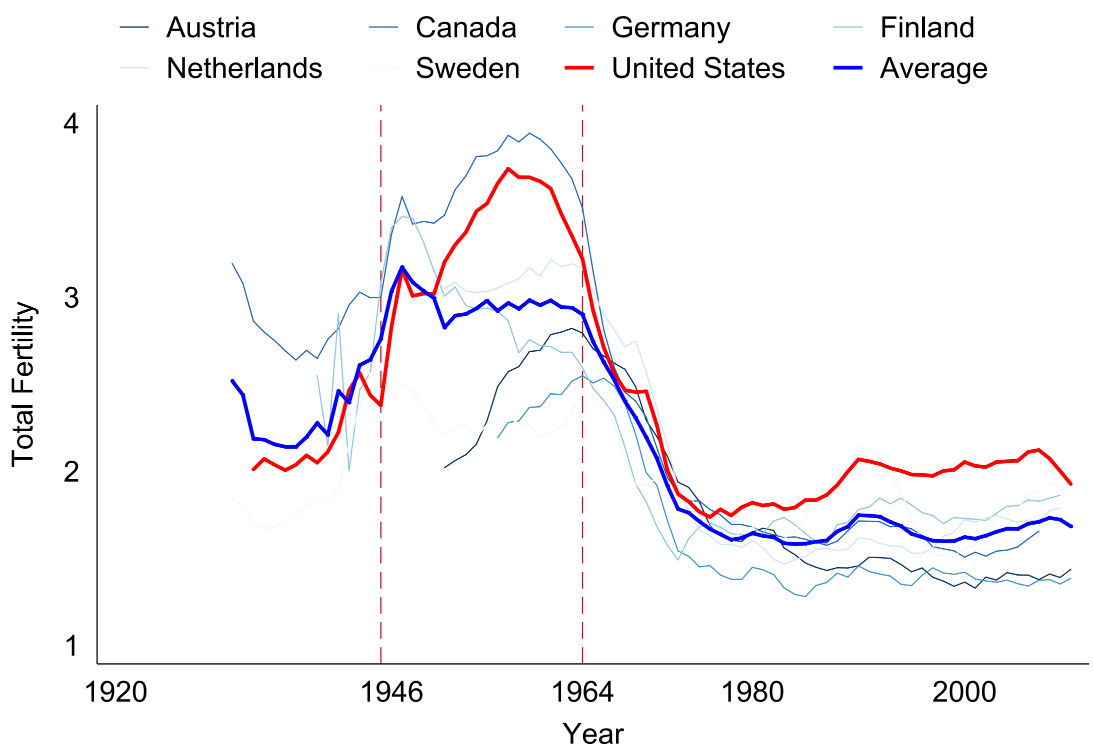
We find that this return *ceteris paribus* will increase the natural rate of inequality for most countries in our sample, and thus may lead to increases in overall inequality.

The Baby Boom generation, for the US commonly considered those born between 1946 and 1964, represented a temporary upwards deviation from developed countries' otherwise stable demographic trajectories. This can be seen in Figure 3.10 which reports long-run fertility data for a selection of countries. A first observation is that the Baby Boom was a common feature across many developed countries.[21] Although, there are variations in timing and magnitude these fail to mask the overall scale of the boom - nearly an extra child per woman for 18 years. Also, notable is the rapidity with which it began and ended. This large, sudden, and in demographic terms brief, rise in fertility has led to a one generation distortion in

[21] All data are from the Human Fertility Database (2013). Germany refers to West Germany only, France excludes the overseas territories. The 'Average' series is the annual arithmetic mean of available observations.

the demographic structure of the affected societies. This shock to the demographic pyramid provides an interesting natural experiment for us to study as the demographics return to their long run steady state following the departure of the Baby Boom generation. Our analysis suggests that recent increases in natural inequality will be permanent, and continue as the share of Baby Boomers in the labour market and overall population declines, with increases of up to 10 percent in inequality as societies return to the demographic steady-state.

Figure 3.10: The Baby Boom



Source: Authors' calculations using data are from the Human Fertility Database (HFD), 2013.
Notes: The y-axis reports the number of children born per woman in a given year. The blue line is the (unweighted) mean fertility rate across the six countries reported. The red line highlights the USA for clarity but is otherwise identical in construction to those for other countries. The dotted vertical lines indicate the beginning and end of the baby-boom.

Future Levels of Inequality

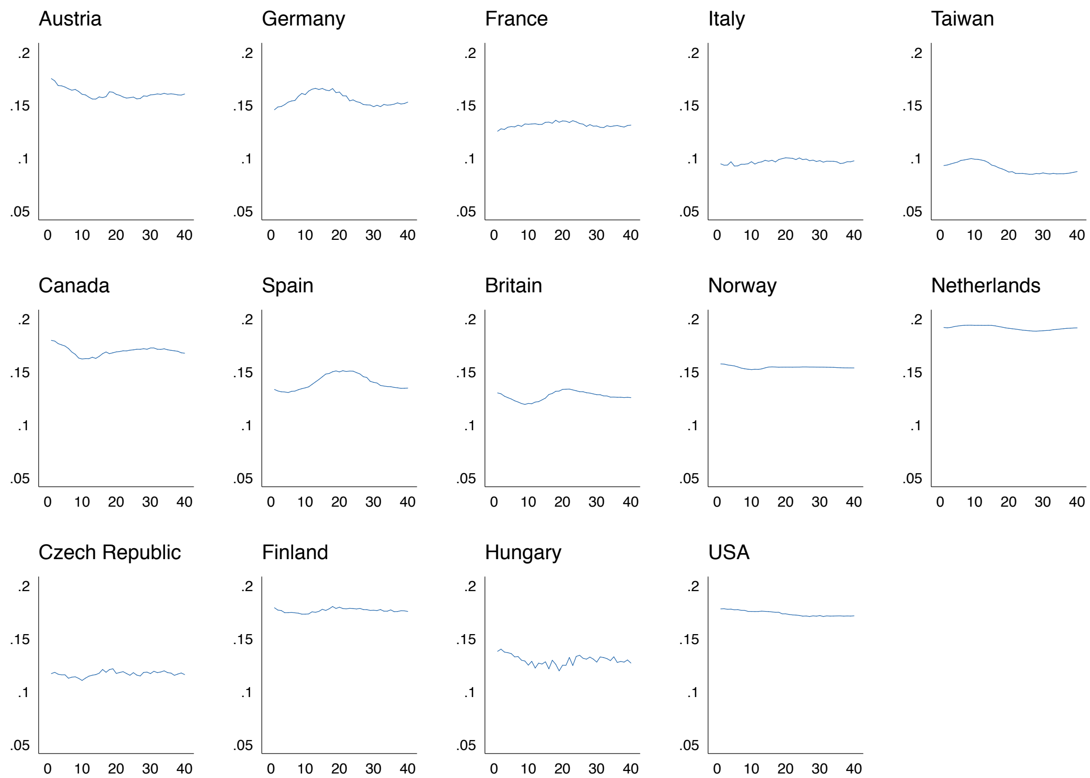
In order to study the impact of the Baby Boomers we simulate future population cohort sizes using age specific data on birth rates, death rates, and population cohort size. We do this using a Leslie matrix, a standard approach in Demography, in which the birth and death rates define a transition matrix that projects the cohort sizes next period given the current sizes.[22] Then, because the natural rate of inequality only requires cohort or age-group specific income shares, we can then use the projected cohort sizes to scale these income shares, giving estimates of natural inequality under the new demographics. This process can be repeated to obtain projected demographics at any given time horizon.

We make two key assumptions for this exercise. Firstly, that the life-cycle earnings profile is stationary. Secondly, we fix the relative size of the working cohort sizes. That is, we assume that the labour market participation and unemployment, and thus inactivity, rates, will remain fixed for each cohort over time. We are asking *ceteris paribus* what will happen to the level of natural rate inequality in a society in the future if all that is going to change is relative cohort sizes. In particular, we can expect to see the society returning to its normal demographic pyramid following the shock of the Baby Boom generation. This assumption entails also not making any inference regarding expected immigration. Thus we are assuming that this will be such that the relative size of the working cohort is fixed.

Thus, for the 15 countries for which suitable fertility and mortality data are available, and are part of the LIS data, we project expected levels of natural labour income inequality. Figure 3.11 plots projected natural inequality for the next forty years. We choose this horizon as by this point the children of the Baby Boomers

[22]The key difference between our approach and standard population forecasts is that, for simplicity, we abstract from anticipated net migration and advances in life-expectancy.

Figure 3.11: Simulated Natural Rates of Income Inequality



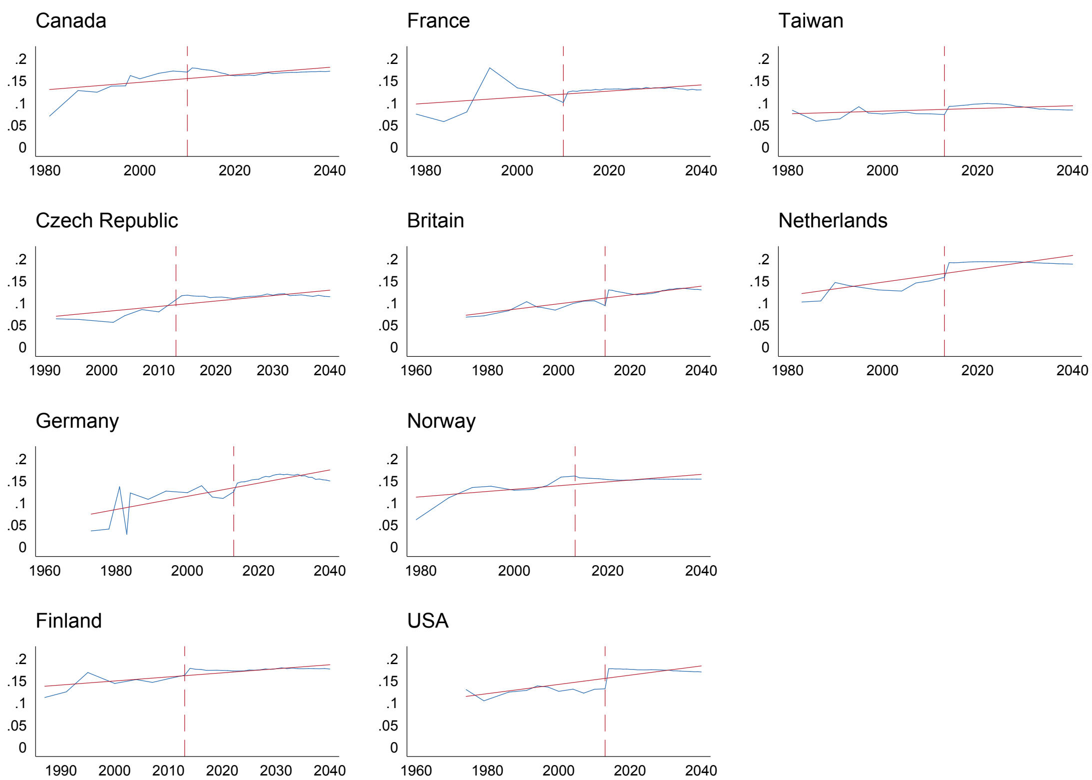
Source: Simulations use data from the Human Mortality Database (2013) and Human Fertility Database (2013) and earnings profiles are taken from the most recent data available in the LIS database.
Notes: On the y-axis is the *Natural* Gini Coefficient and time (years in the future) is on the x-axis. We project the population distribution for up to 40 years in the future by which time all societies will be extremely close to their steady state.

have largely left the labour market and so the population will be approaching its steady-state. The key prediction is that in almost all countries natural inequality will remain at its current level at least. A second prediction is that natural inequality will be much less volatile than in the past, although other than in the United States and Norway it will continue to fluctuate. Both of these results are consistent with our intuitions, as the Baby Boomers either have now retired or will do in the next few years. Seemingly, in the past the presence of the Baby Boomers reduced natural inequality, offsetting and thus masking increases in adjusted inequality. Any future rises in adjusted inequality will translate directly into increased overall inequality.

A second prediction concerns the timing of the fluctuations, which are expected to be largest around twenty years from now, when mortality rates for the Baby Boomers will be highest. This effect seems particularly pronounced for France, Germany, Spain and Britain. To further look at how these projections compare with the historical data, we plot them together in Figure 3.12 along with the line of best fit from a linear least-squares regression in red.[23] The vertical red dashed line represents the point at which the simulation starts. To the left of this line are the historical results from LIS, and points to the right are the projected levels of inequality. Taken together it seems that future increases in natural inequality would represent a continuation of the historical trend. Historically, this presumably reflects the increased numbers of older people in the population due to improved health, and it is important to note that any continued improvements will likely increase natural inequality further. Most countries are forecast to experience a five to ten percentage points increase in the natural rate relative to the 1980's by the 2040's. This suggests that in the absence of more migration or changes in fertility patterns that there is unlikely to be any reduction in natural inequality, to offset trends in excess inequality, in the foreseeable future.

[23]The reduced set of countries reflects data availability.

Figure 3.12: Historical and Simulated Future Rates of Income Inequality



Source: Simulations use data from the Human Mortality Database (2013) and Human Fertility Database (2013). Historical data are taken from the LIS, the Earnings profiles for the projections are taken from the final wave of the LIS.
Notes: On the y-axis is the *Natural* Gini Coefficient and the x-axis plots the year. The dashed vertical red line signals the end of the historical LIS results and the beginning of the projected trend. The solid red line is the least-squares line of best fit for the entire time period.

3.4 Conclusion

Even a society in which everybody is the same at the same stage of the life-cycle will exhibit a substantial degree of income and wealth inequality. In this paper we take this notion to the data in order to quantify the share of observed income and wealth inequality that is attributable to life-cycle profiles of income and wealth. The data reveal that natural inequality is a substantial component of actual inequality. Treating the natural rate as the benchmark, and thus analysing excess or adjusted inequality suggests that recent increases in income inequality in the US are both larger than the actual rate would suggest, and represent a distinct change from the period pre-1990. It is also clear that natural inequality is of first-order importance in understanding variation in other developed countries and the variation between them. A similar analysis for wealth inequality suggests that natural inequality is also important to understand trends in wealth inequality, although it accounts for a smaller component of actual wealth inequality. Allowing for differences in natural inequality suggests the USA is much less of an outlier compared to other countries. To home in on the role of the demographic structure for inequality we close our analysis by focusing on the impact of the bulge on the demographic pyramid generated by the Baby Boom generation. This shows that the as cohort shares transition back into their long-run equilibrium levels, natural inequalities of income will fluctuate and reach a new higher level of steady state natural rate inequality.

Chapter 4

The Declining Fortunes of the American Worker

4.1 Introduction

Many of us take economic progress for granted. We both expect that normally our earnings will increase from one year to the next, and that we should be more prosperous than our parents and grandparents were. Yet, this expectation is increasingly misaligned with recent experience. U.S. real Median Earnings have seen little improvement since 1980.[1] Meanwhile, US GDP per capita has nearly doubled since 1980.[2] This increase reflects both growth in women's market earnings due to greater labour market participation as well as increases in the earnings of the richest.

Focusing on men alone the picture is starker still. In 1965, in real terms, the median American man was earning $31,734, and the income of the tenth percentile

[1] See Figure C.1 in the Appendix.

[2] Measured in 2010 Dollars, it was $28,589 in 1980 and $54,551 in 2018. Data from: `https://data.worldbank.org/indicator/NY.GDP.PCAP.KD?locations=US`

was \$10,578. By 1995, the figures were, \$30,577 and \$8,197. Thus, the average American man selected at random was poorer 30 years later. Fifty years later, in 2015, they were still no better off than in 1965 with incomes of \$31,635 and \$8,436 respectively.

This paper studies this stagnation from an inter-generational perspective. We trace the real earnings of each generation over the lifecycle and document that for each generation subsequent to the Baby Boomers, living standards have declined substantially in real terms. That is, rather than being richer than their parents, the median member of *Generation X*, born between 1965 and 1979 or *Millennial* born in the period 1980–1999 is poorer at every point during their working lives than their parents were as members of either the *Boomer* or *Silent* generations, born 1946–1964 or 1925–1945 respectively.

We can see this in Figure 4.1 which plots the median earnings at each age for white, male, high-school educated, Americans by decade of birth. Comparing the median wage across cohorts, we can see that those born in the 1920s, entering the labour market between 1940 and 1950 were earning over \$30,000 in 1999 dollars by their early 30s when we first observe them. This is less than those born in the 30s and the 40s, but interestingly this cohort have the highest peak earnings of any cohort, at around \$45,000 in 1999 dollars. The curve for the 1930s cohort is broadly similar. However, the 1940s cohort saw their wages drop by nearly a quarter in real terms at the beginning of the 1980s and never recover. A similar change seems to affect the previous cohorts, but later in life where it is conflated with retirement. Cohorts from the 1950s on see comparatively little wage growth, earning less at every point in their lives than their forebears. The difference is quantitatively large, a white male high-school graduate born in the 1930s is earning about \$40,000 by age 40, their son, say born in the 1950s, makes $\$30,000$, their grandson born in the 1970s had a median wage around $\$25,000$.

This purpose of this paper is to study this phenomenon of inter-generational declines in median wages. As such it is related to and builds on the recent work of Guvenen et al. (2017), discussed in detail below, who are the first to our knowledge to document this phenomenon. The first section of this paper, discusses inter-generational changes in earnings providing evidence of similar patterns, with few exceptions, across demographic groups, and the population as a whole. In particular, we find that the median earnings of male college graduates have also declined. This is also true for women at least since the Boomers, and African-American and Hispanic-American men. The key exception has been the substantial improvement in the earnings of African-American and Hispanic-American women. Both graphical analysis at the cohort level, as well individual level regression estimates show that this is a general phenomenon. Considering conditional demographic and educational controls, as well as state and industry fixed-effects we find we find that Boomers earned 4% less than the Silents. With Gen. X'ers and Millennials earning 8% and 16% respectively. We also show that as well as earnings being lower, that later generations have had to wait longer to attain peak earnings.

We show that a reduction in hours worked between generations cannot explain these declines, and given that overall U.S. has grown strongly throughout the period we study (Jorgenson et al., 2008, Bloom et al., 2012) another explanation is necessary. Thus, the second part of this paper builds on the recent literature documenting and analysing the declining U.S. labour share to argue that inter-generational differences in the labour share are, in part, responsible for declining US real wages. We find evidence that the labour share is indeed 6% lower for Millennials compared to Boomers.

We next perform a variance decomposition analysis which shows that while generation cannot explain much of the variation in log wages after accounting for changes in the distribution of wages over states and sectors and education and demographic characteristics that it is still an important determinant of the labour share. We argue that this implies that decline in real wages is driven, in part, by reductions in the labour share.

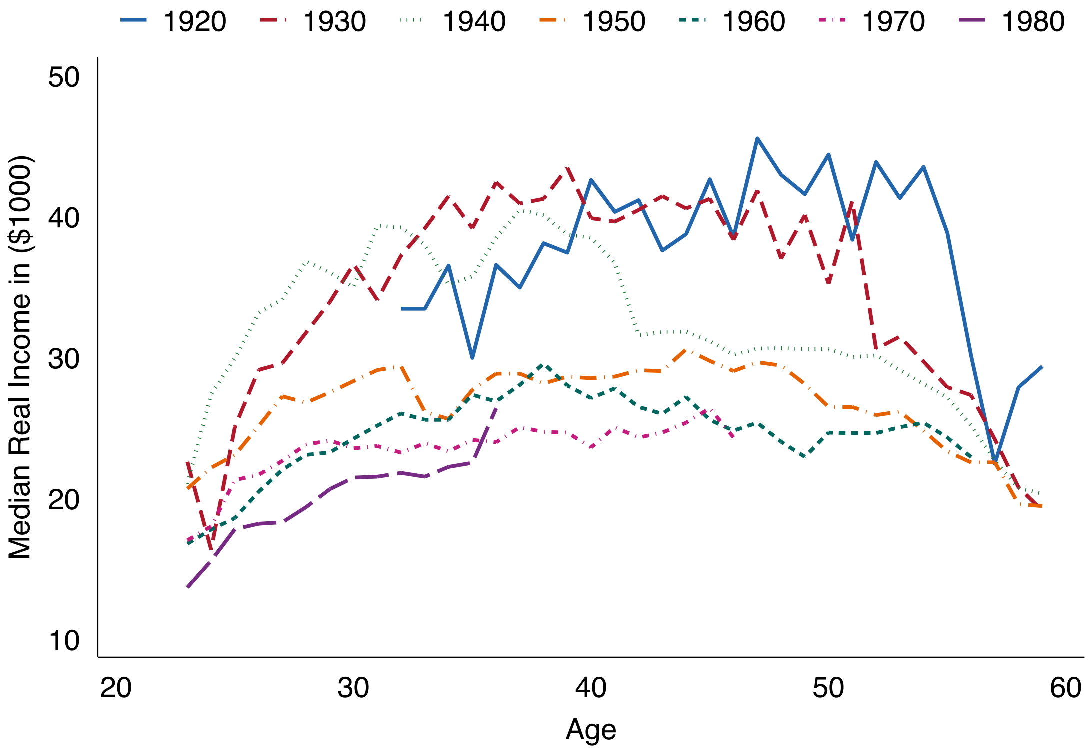
A potential concern is that our focus on real wages means that we do not adequately capture improvements in the quality of goods and services, or improvements in working conditions. In the final part of the paper we study this argument quantitatively asking what level of hedonic improvement would be necessary to compensate for the observed differences in income. We begin by assuming log-utility for tractability and focusing only on differences in earnings. This suggests the relative life-time quality level would need to be around 20% higher for those born in the 1970s compared to those born in the 1920s. We consider an iso-elastic utility function that also includes the disutility of work as in Mankiw and Weinzierl (2006), to analyse the limitations of this approach.

Our work relates to several strands of literature. Firstly, the literature on 'progress'. Secondly, it relates to the literature on the labour share. Thirdly, on inequality and mobility, Finally, it is connected with the literature on secular stagnation.

This paper is part of a nascent literature that studies the comparative fortunes of different cohorts. There are two recent papers that also take a cohort approach to lifecycle earnings and similarly document the decline in the incomes of younger generations.[3] To our knowledge Guvenen et al. (2017) is the first to provide evidence of this. They use an extract from the U.S. Social Security Administration register

[3] Acemoglu and Autor (2011) document the declining income of lower skilled workers but not the disparity across generations.

Figure 4.1: Median Wage of White Male High-school Graduates, by Decade of Birth, Over the Lifecycle



Source: Current Population Survey (CPS)
Notes: Includes the total population, wages are adjusted for inflation, and individual weights are used.

data to show that total life time income is at least 10% lower for the most recent cohort they study (those entering the labour market in 1983) compared to the first (those entering in 1957). Importantly, they are able to rule out a substantial effect of non-pay benefits such as employer provided health insurance and pensions, and the choice of price deflator. An important second contribution of their paper is to study the evolution of gender differences in incomes between men and women across cohorts. The life-time earnings of the median women have converged towards that of the median man, from 37% for the 1957 cohort, to 60% for the 1983 cohort, with this being partially driven an increase in the number of years worked by younger women (aged 25–34) but mainly by increasing incomes of those women in work. Interestingly, while this reduction in earnings differences between men and women has reduced within cohort inequality, this reduction has been offset by an increase in income inequality within each gender group. Moreover, while an analysis of more recent cohorts suggested continued growth in women's earnings until 1998, they have subsequently stagnated – although, as they caution, this may reflect the consequences of the Financial Crisis. Finally, they adopt a slightly different perspective and analyse the distribution by decile and gender of the total income of each cohort. They show that while women's share has increased for every decile that this increase has been larger for the highest earning, while only those in the top few percentiles of the male life-time earnings distribution have seen an increase, with most experiencing a substantial decline.

Relative to Guvenen et al. (2017) the contribution of this paper is to compliment their analysis by using CPS data (and matching it to US Economic Census Data) to both examine differences across and within levels of educational achievement as well as differences in the experiences of different racial groups in addition to comparing men and women. Moreover, we seek to understand the drivers of these patterns in terms of the changing geographic and sectoral distribution of economic activity, the declining labour share of income, and skill biased technological change. Further, whilst they provide an important discussion of the role of alternative deflators and non-pay benefits, we analyse the role of hedonic progress, etc.

Borella et al. (2019) also take a generational approach and note that as well as Americans born in the 1960s having lower earnings than those born in the 1940s, that they have lower life expectancies and higher out of pocket medical expenses. Using the model introduced in Borella et al. (2017), which incorporates the behaviour of both single people and couples, to calculate the effects of these three changes and find that they are equivalent to an asset transfer of 12.5% of their lifetime earnings or $\$126,000$ aged 25. For single women the figures are $\$44,000$ (7.2%) and $\$132,000$ (8%) respectively. The aspect of this paper that is closest to what they do is Section 4.5 where we try to quantify the degree of hedonic progress that would equalise the value of consumption minus the disutility of work across generations.

Secondly, our focus on the labour share has much in common with an important recent literature that highlights the declining labour share. The paper closest to this one is Autor and Salomons (2018) who provide evidence that automation has not led to reduced employment but can partly explain the decline in the labour share. They build on the framework introduced by Acemoglu and Restrepo (2018) which while predicting that while automation may reduce both wages and the labour share, along the balanced growth path that this process is self-limiting as automation increases the productivity of labour and creates new tasks. Automation is only one of several secular trends that have been found to exert downwards pressure on the labour share. Autor et al. (2017b) argue that another possible cause of the global decline in labour share is more productive firms (because of either globalisation or technological change) gaining higher concentration in their respective industries, creating so-called *superstar firms*. These firms have a higher mark-up and lower share of labour in sales and value added than compared with previous times. As a result, these firms are able to operate across multiple industries and so the within industry labour share will fall. Karabarbounis and Neiman (2014) show that the labour share of income is falling in most countries and attribute this decline to falls in the price of investment goods. Grossman et al. (2018) show that the slowdown in productivity growth in recent years may lead to a lower labour share. Finally, Elsby et al. (2013) consider the roles of two other key recent trends: Unionisation and off-shoring. They find little role for unionisation but emphasise the impact of off-shoring.

While we focus on the absolute standard of living, Piketty and Saez (2014) links the declining labour share to increasing levels of inequality (Piketty and Saez, 2003b, Saez and Zucman, 2016). Similarly, our results can be thought of in terms of the growing consequences of the lottery of birth documented by Chetty et al. (2014b): the average ticket is now a losing one in absolute terms. Others have highlighted non-economic consequences of such trends. For example, our findings might be read through the lens of the argument of Friedman (2005), in which the lack of broad-based economic progress can imperil the moral quality of society, particularly his argument that a lack of economic progress may imperil the quality of democracy. Likewise, Case and Deaton (2015) document increasing morbidity amongst those aged 45–54, so called 'deaths of despair'.

This paper is organised as follows. The next section describes life-cycle earnings profiles for different generations and sub-populations. It also discusses the role of hours worked. Section 4.3 provides evidence from individual level regression analyses and explores the role of the declines in the labour share. Section 4.4 reports the results of a variance decomposition analysis. Section 4.5 considers how much hedonic improvement in consumption would be necessary to equalise consumption levels across generations. Section 4.6 briefly concludes.

4.2 Life-Cycle Earnings

We will rely on the Current Population Survey (CPS) for the bulk of our analysis, as well as the Economic Census. Further discussion of these data and how we handle them may be found in Appendix C.1.

4.2.1 Wages

Life-time wages

In this section we report similar results to Guvenen et al. (2017) but take the opportunity of the detailed demographic data in the CPS to decompose by education and race as well as by gender like they do. We also work with a slightly different sample. Guvenen et al. (2017) mostly focus on those cohorts who were 25 in 1957 to those who were 25 in 1983, such that they can follow each cohort for the ages 25-55, and define life-time earnings over that period. This has the important advantage that education is largely complete by age 25, and relatively few people retire or otherwise leave the workforce before age 55. We choose to focus on a longer time span – on those born between 1925 and 1999, at the price of not being able to follow the last generation, the Millenials throughout their lives. However, by now the oldest of these are nearly 40 and thus we are in position to compare their comparative fortunes to this point. This comparison is worthwhile because by around this age many Americans would hope to have bought a house, started a family, etc. More generally, a standard discounting argument implies that an income profile that offers greater earnings early in one's life, holding total earnings constant, is to be preferred. Thus, the shape as well as the level of each generation's earnings profile matters for welfare, and may be usefully compared by age 40.

Table 4.1: Different Birth Cohorts.

2000 – Present	Generation Z
1980 – 1999	Millennial's (Gen. Y)
1965 – 1979	Generation X
1946 – 1964	Baby Boomer's
1925 – 1945	Silent Generation

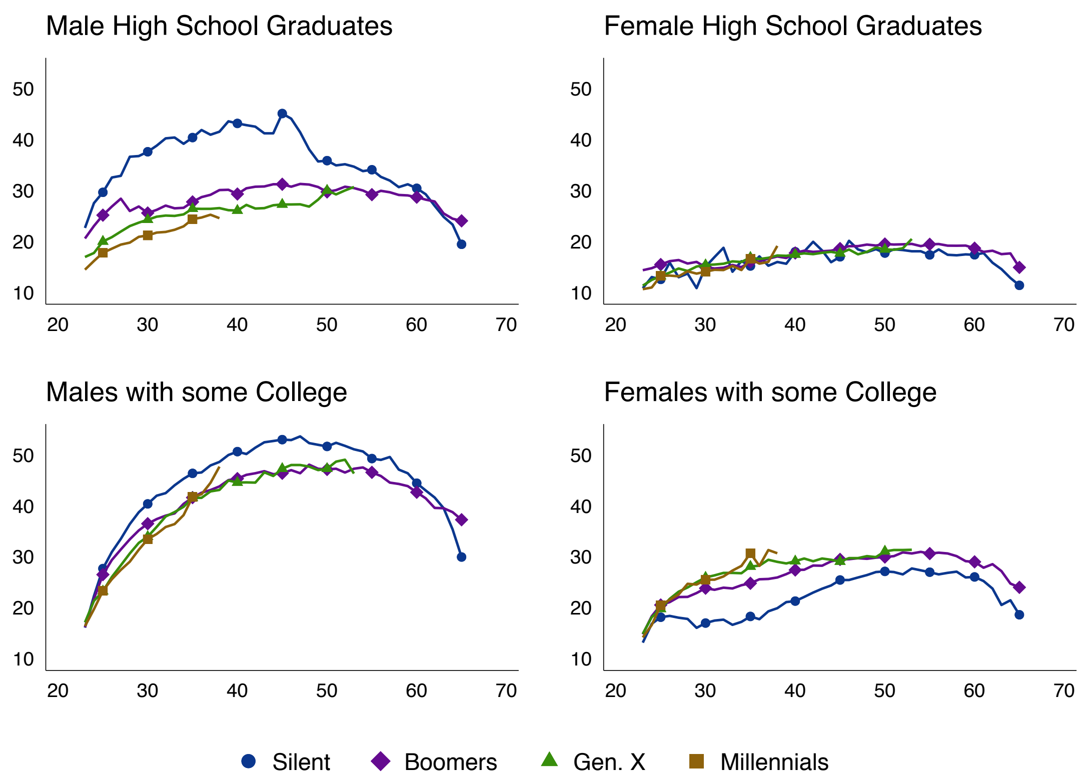
For exposition purposes, we divide the Americans in our data into generations, as they are typically defined. Table 4.1 displays how we define each generation. The top-left panel of Figure 4.2 is analogous to Figure 4.1 except now the median wage over the lifecycle is plotted by generation. We can see clearly that the Silent generation (born 25-45 denoted by blue circles) have higher earnings at every point than the Boomers (purple diamonds), Gen. X'ers (green triangles) and Millennials (brown squares). Moreover, this difference is substantial, nearly $\$20,000$ a year, at age 45 or two thirds of Boomer earnings. While the Baby-Boomers earn less than the Silent generation, they do earn more than the later two generations. Moreover, they hit peak earnings sooner, by their late 20s, while Gen. X'ers experienced a much slower growth in their earnings, even if they seem to have converged by age 50. This is also true for Millennials. Figure 4.3 reports the same data but now with year rather than age on the x-axis.[4] This makes clear the declining fortunes of High School graduates. Each generation's curve is below (excluding a drop off in earnings for the Silent generation from age 50 onwards) that of the one before. The average across all generations, not plotted, thus declines as Gen .X'ers and Millennials start to replace Silents and Boomers. Note, that we might expect, given substantial economic growth, the opposite: That each generation would start from a higher point that the preceding one and increase from there such that the curves would intersect.

The bottom-left panel presents results for men with at least some College education. We see that, again, the median wage of the Silent generation is higher at all points in their career. This means that the decline of real wages is not limited to High School Graduates, suggesting that the phenomenon is not limited to those in lower-skill occupations. But, the difference with the Boomers is smaller now, and

[4]Note, that there will be some difference in the estimates since Figure 4.2 takes the median of all members of a given generation at a given age. While, Figure 4.3 reports the median in a given year of all members of a generation, who will hence be of a range of ages.

there is no appreciable difference between the Boomers and the subsequent generations. This is consistent with skills-biased technological change advantaging those with more formal education in subsequent generations relative to those with less in their generation and thus reducing the gap between generations of the more educated. This explanation is also consistent with what seems to be some evidence of improved earnings for Millennials who attended college in the last couple of years relative to the Boomers and Gen. X'ers. But, without more data it is not possible to rule out that this is just a short-term fluctuation.

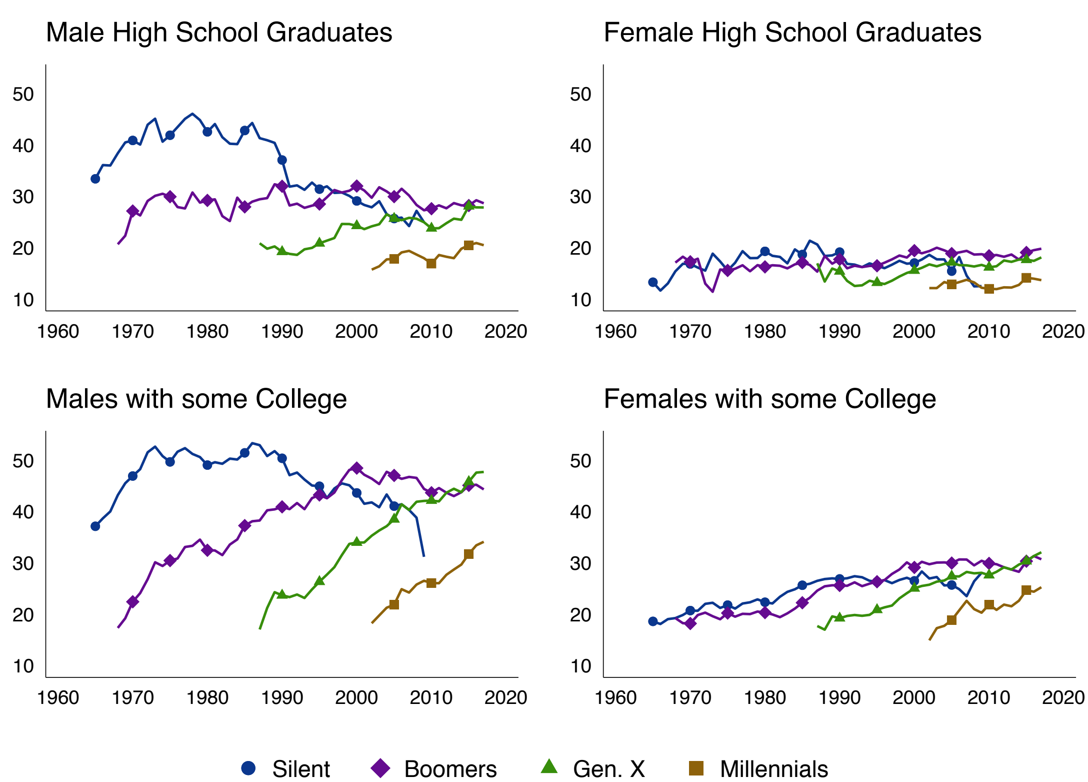
Figure 4.2: Median Wage by Generation over the Lifecycle



Source: Current Population Survey (CPS)
Notes: Includes the total population, wages are adjusted for inflation, and individual weights are used. 'some College' includes those who attended college but didn't graduate, those did graduate, and those who have an advanced degree.

Looking at the bottom-left panel of Figure 4.3 again reinforces the point. We again can see lower earnings at every point for each subsequent generation and more notably for this sample, pronounced generational differences in the rate of progress over the lifecycle. This can be seen by comparing the difference between the Boomer's curve and that of Gen. X or the Millenials, which are substantially flatter at the beginning.

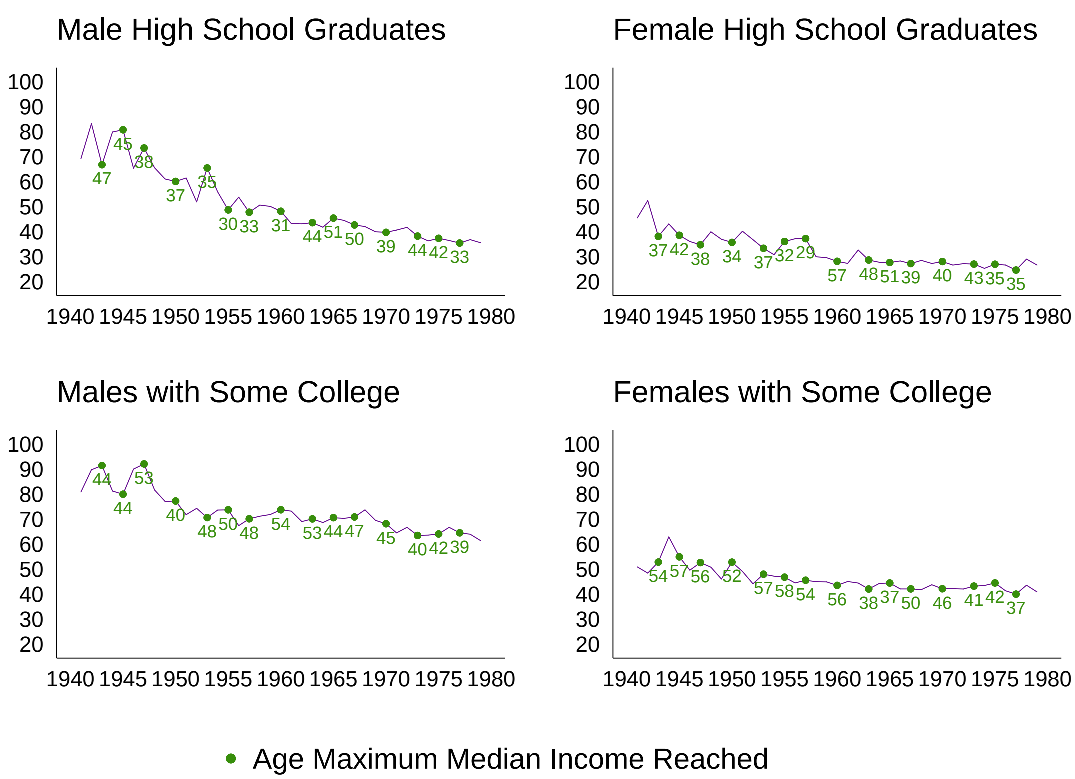
Figure 4.3: Median Wage (in $1000) for each Generation over time



Source: Current Population Survey (CPS)

The blue line in the top-left panel of Figure 4.4 plots the maximum median wage reached by year born for male high school graduates. We then annotate these points (in green) with the age when this income was reached. We see that median American born in 1945 had maximum earnings of just over $\$65,000$ which they achieved age 47. In comparison, the maximum median wage of those born 10–15 years later was substantially, but was achieved by their early 30s. Those born from around 1961 to 1970 not only had lower maximum earnings but they did not receive them until they were 50. More recent cohorts had again lower maximum earnings, albeit slightly earlier at ages 42-44. Given the effects of the Financial Crisis, it maybe premature to reach a conclusion about those born in the late 1970s as it is conceivable that their earnings will still increase meaningfully.

Figure 4.4: Maximum Median Wage by Year Born



Source: Current Population Survey (CPS)
Notes: Includes the total population, wages are adjusted for inflation, and individual weights are used.

By Gender

The right two panels of Figure 4.2 show the results of the same analysis for Women. Looking first at the results for High School Graduates in the top-right panel it is clear that women's median earnings are on average, across the life-cycle, and across all generations considerably lower than those of men. It is also clear that there is little progress across generations. This result is in contrast to the findings of Guvenen et al. (2017) and this may reflect differences in the origins of the data used and the sample definition. One appealing feature of Guvenen et al. (2017) is that they are able to use administrative data providing recorded rather than self-reported earnings data. A disadvantage of this is that it may exclude unrecorded earnings, which our data should capture. Looking at the top right panel Figure 4.3 we see that each generation seems to converge to within a few thousand dollars of a median of $\$20,000$.

The bottom-right panel now shows the results for women who attended college. Here, the opposite story is true. Each generation seems to be out earning the one before it. Thus, the Silent generation is now at the bottom followed by the Boomers, the Gen. X'ers and finally Millennials. Consistent with this, in Figure 4.3 we now see this pattern of the median earnings of each cohort intersecting with those before it (albeit not yet for Millennials). This suggests, that perhaps the growth in women's earnings documented by Guvenen et al. (2017) are due solely to the growth in the earnings of college educated women and the growth in the proportion of women attending college.[5]

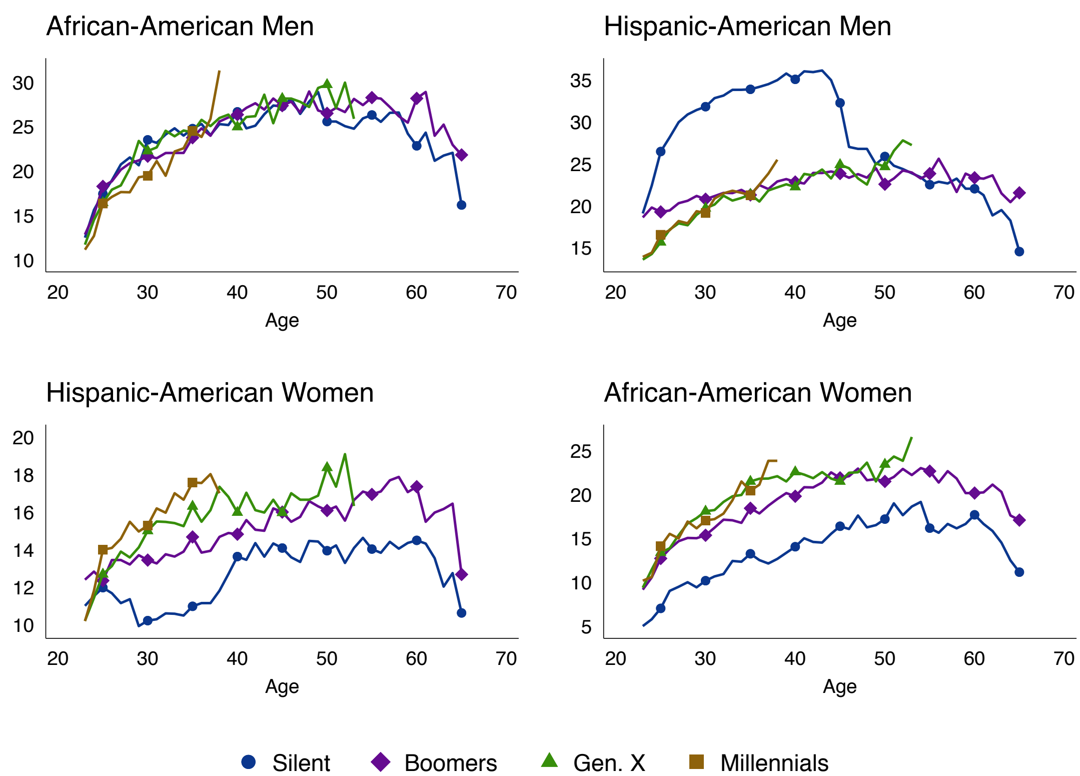
[5] Guvenen et al. (2017) restrict the sample to those with consistent labour market engagement and a minimal level of income that may disproportionately exclude less-educated women, who may be more likely to be in informal employment.

By Race

Another key margin of income inequality is race: African Americans and Hispanic Americans continue to have lower average incomes than other Americans (Fryer, 2011). Given that around the beginning of our sample period, the passage of the Civil Rights Act made discrimination on the basis of race illegal, and recent evidence suggests that discrimination can account for a relatively small proportion of the racial earnings gap (Fryer, 2011). Thus, we might expect subsequent generations of Black and Hispanic men to have higher incomes than those of the Silent generation even if male earnings in general are declining. Similarly, we expect more rapid growth in the earnings of Black and Hispanic women. However, inspection of the top two panels of Figure 4.5 which reproduces Figure 4.2 for Black and Hispanic men suggests that this is not the case. Incomes at each point in the life-cycle are broadly constant across all four generations of Black men. It is unclear why relative pay of Hispanic Silents was so much higher than subsequent generations, but focusing on the Boomers onwards we see no evidence of an increase in the wages of Hispanic men either, and indeed arguably a decline. Of course, migration makes comparisons across generations more difficult and it maybe that the lack of earnings growth is due to a composition effect. This would explain, potentially, the substantial decline in earnings from the Silent Generation to subsequent generations.

The bottom two panels report results for Black and Hispanic women respectively. Now, we see clear signs of increasing incomes from one generation to the next. Looking first at the evidence for the African American women in the bottom left panel we see that working women of the Silent Generation earned around $\$5,000$ less than Baby Boomers. Who in turn earned less, albeit not as much, less than Gen. X'ers and Millennials. A similar, but arguably more pronounced pattern can be seen in the bottom right panel for Hispanic women. Now, as well as daylight between the Silents and the Boomers there is a clear difference between Boomers and Gen. X'ers and in turn them and Millennials. Common to both Black and Hispanic women is that Gen. X'ers and particularly Millennials, both show signs of rapid income growth during their 20s and 30s. This is consistent with the closing of the gap in college enrolment rates in both populations compared to American women as a whole.

Figure 4.5: Median Wage by Generation over the Lifecycle



Source: Current Population Survey (CPS)

4.2.2 Hours Worked

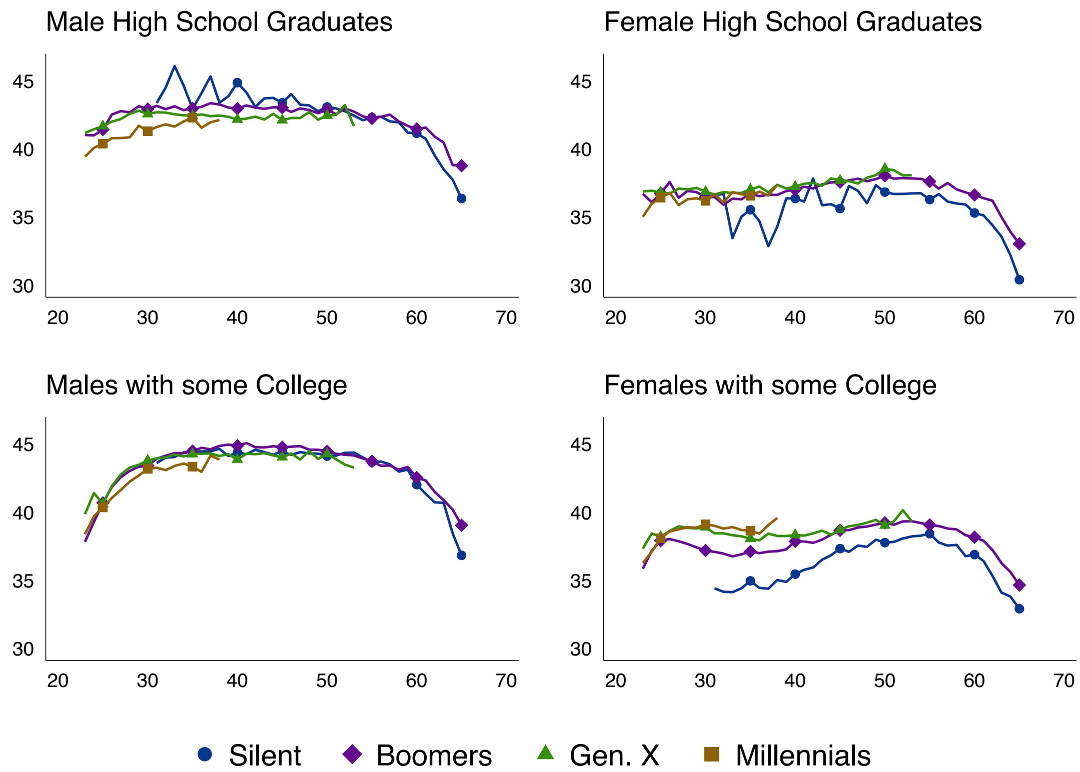
One possibility is that stagnant earnings reflect in part reductions in hours worked. This alters the comparison across generations since we normally presume that welfare is decreasing in hours worked. Figure 4.6 reports the number of hours usually worked per week over the life course for each generation. Looking at the plots for men in the left column we see that, consistent with existing evidence (Blundell et al., 2011, McDaniel, 2011), that there have been no abrupt changes in the number of hours worked. There is some evidence that Silent High School graduates worked more on average and particularly in their 30s, and that Millennials seem to work less than Boomers and Gen. X'ers, but the overall differences are relatively small. There are, as expected, greater changes for women. With a clear increase in hours worked by all generations subsequent to the Silents for all women. As well as smaller, but still noticeable, differences for college educated women between the Boomers and Gen. X'ers (and Millennials). Figure C.6 in the Appendix provides analogous plots for African-Americans and Hispanic-Americans showing similar patterns. Overall, it seems reasonable to conclude that there has not been a sufficiently large decrease worked by American men to imply a rising real hourly wage.

4.3 Individual Level Data

The preceding graphical analysis makes clear that later generations of American men have to date suffered decreasing incomes compared to their elder peers. This is also true for female high-school graduates. But, not for African American or Hispanic women or women who attended college. We now dig deeper, working with individual level data so that we can understand inter generational differences controlling for a range of other determinants.

An important limitation of our analysis is that it is not causal, and we work with a repeated cross-section unlike Guvenen et al. (2017) who use panel data and are thus able to control for a broader range of factors. The compensation for this limit-

Figure 4.6: Hours Worked by Generation over the Lifecycle



Source: Current Population Survey (CPS)

ation is that the CPS are relatively rich, and in this section we are able to investigate how these patterns identified in section 4.2 reflect structural change in the U.S. economy. Specifically, whether or not there remains inter-generational differences in incomes once we allow for the changing sectoral composition of the U.S. labour market, the changing geographical distribution of economic activity, or the increasing returns to education.[6] It also means we are able to build on the recent literature documenting and explaining the decline in the labour share (Karabarbounis and Neiman, 2014, Elsby et al., 2013, Piketty and Saez, 2014, Autor et al., 2017b, Grossman et al., 2018) and particularly Autor et al. (2017b) to test whether, other

[6]Acemoglu and Autor (2011) provide a detailed discussion of the leading models/data.

things equal, the decline in median life-time earnings is, in part, due to reductions across generations in their average labour share.

This section proceeds as follows. Below we outline our measure of the labour share and then introduce our regression specification. We then present our estimates of inter-generational differences in log wages and the labour share.

4.3.1 Calculating the Labour Share

We compute the labour share for each generation, in each year, and in each industry. We do this using a similar approach to that of Autor et al. (2017b). Like them we define the labour share of firm k, λ_k, as the ratio of annual payroll to the firm's total value added. We also use their cross-walks to combine the CPS and Economic Census. Details are provided in Appendix C.1.

We define the labour share in industry i as the size weighted average of λ for all firms in that industry using the data from the BEA. Linking this with data on demographic information by industry from the CPS means we can then compute the labour share of a given generation as the employment weighted average of industry labour shares. We assign firms to industries on the basis of their 1-digit NAICS codes. The details of how we merge the CPS data with the U.S. Economic Census are in Appendix C.1. More precisely, we compute the labour share of a given generation g in a given industry i in a given state s in a given year t,

Then, from the CPS we know $w^{st}_{g,i}$ the share of a given generation working in industry i in a given year and state. Having calculated the labour share in each industry in that year and state we then compute the labour share of a generation as the share weighted average. That is,

$$\lambda_g^{s,t} = \sum_i \lambda_i^{s,t} \cdot w_{g,i}^{s,t} \quad \text{such that } w_{g,i}^{s,t} = \frac{n_{g,i}^{s,t}}{N_i^{s,t}} \text{ and } \sum_g w_{g,i}^{s,t} = 1 \tag{4.1}$$

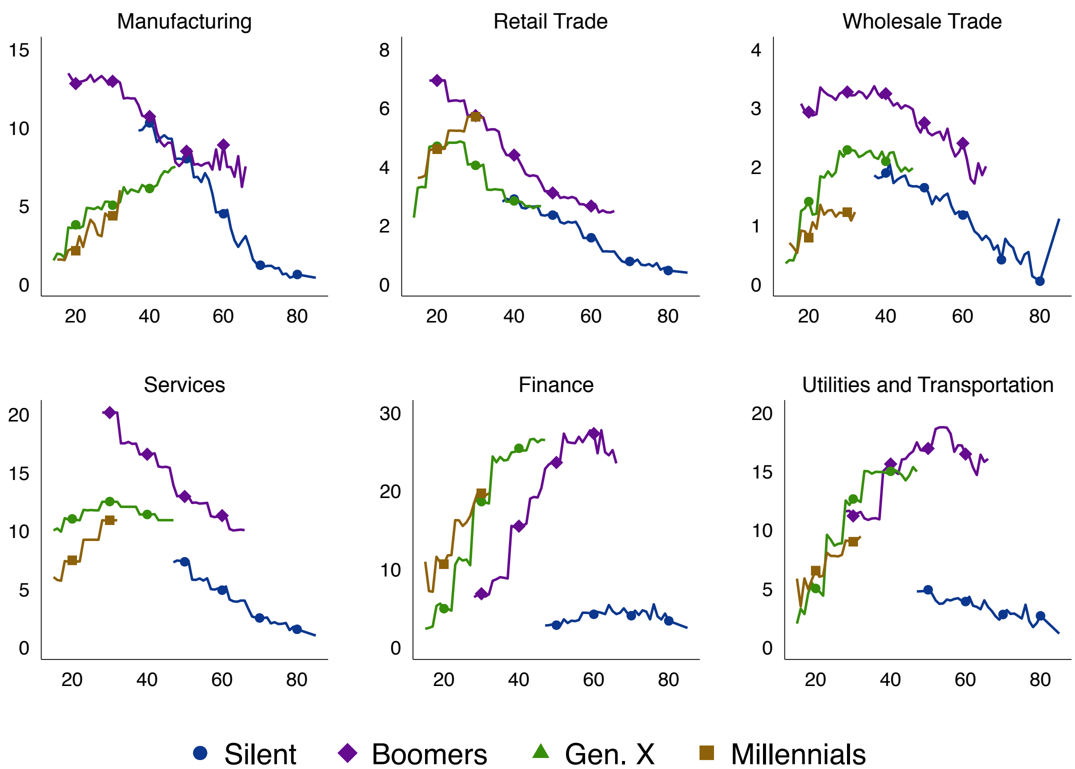
Where the total number of worker in industry i is denoted N_i and the number of workers in industry i which belong to cohort c is denoted $n_{c,i}$. We can also use variation in sectoral composition across states to compute a generation × industry × year specific labour share.

$$\lambda_{g,i}^t = \sum_t \lambda_i^{s,t} w_{g,i}^{s,t} \tag{4.2}$$

Figure 4.7 plots the labour share over the life-cycle by generation for each of the 1-digit NAICS. Looking at the data it is clear that the Boomers have experienced consistently higher labour share at each age than any other generation. This is an interesting contrast to the previous results for income which had the silent generation earning more. One explanation, discussed in detail by Guvenen et al. (2017), is that this could reflect changes in non-pay costs such as health insurance. It could also reflect changes in the number of hours worked although the results in section 4.2.2 suggest the changes have been too small to account for all of the change. Equally, it might reflect changes in the denominator and thus changes in the average, and distribution of, firms' value-added as argued by Autor et al. (2017b).

Also interesting is the variation across industries, not only does the life-cycle average vary considerably across industries, but the trajectories over the lifecycle are also quite different. In some industries such as Services, Finance, or Manufacturing there seems to be an initial upswing in the labour share, perhaps reflecting increased bargaining power as specific skills are obtained and labour markets become thinner. Whereas, in the Retail Trade, the labour share consistently drops over the life-course.

Figure 4.7: Labour Share by Age for each Industry and Generation



Source: U.S. Economic Census & CPS

4.3.2 Regression Analysis

We work with the following specification where $c_{j,t}$ is either (log) wages or the labour share of individual j in a given year.

$$c^i_{j,t} = \gamma^i_g + X'_{jt}\beta^i + \delta^i_t + \delta^i_s + \varepsilon^i_{j,t} \tag{4.3}$$

We estimate this equation separately for each industry i so that we allow both the structure of earnings differences across generations to vary by industry. To allow for these changes in the sectoral and geographic composition of the US economy as well as the changing role of education and demographic characteristics means that we will estimate separate specifications for each industry. At the cost of additional coefficients to consider, this allows for unrestricted heterogeneity across specifications as well as ease of interpretation. It will also highlight important sources of heterogeneity. This also means that we allow the effects of both observables such as age, education, or race and the fixed-effects to vary across industries in an unrestricted way. This is important since, for example, there is good to reason to believe that the return to education may differ by sector. Moreover, as discussed above both theory and prior empirical evidence suggests that we should expect heterogeneity across sectors due both to the potential for automation and the adoption of computers (Acemoglu and Restrepo, 2018, Autor and Salomons, 2018, Burstein et al., 2019) as well as the differential effects of off-shoring and sector specific productivity trends (Elsby et al., 2013, Grossman et al., 2018).

We are most interested in the vector of generational dummies γ_g^i, which capture how the average earnings of each generation differ (with respect to the Silent Generation). β^i captures the effects of a standard set of educational and demographic controls. Specifically, we include a quadratic in age, dummies for being African-American or Hispanic as well as Female and whether the respondent graduated high school or attended college. We control for unobserved sources of local and temporary variation by including a set of state and year fixed effects, d_s^i and δ_t^i.

To build our intuition, Table 4.2 reports the results of a minimal specification in which we omit the controls and the fixed effects from equation (4.3). Looking across the first row we can see that Boomers have, unconditionally, earned more in

Table 4.2: Generational Differences in Real Wages by Industry: Unconditional Results

	(1) *Retail* β / SE	(2) *Wholesale* β / SE	(3) *Services* β / SE	(4) *Finance* β / SE	(5) *Utilities* β / SE	(6) *Manufacturing* β / SE	(7) *Construction* β / SE	(8) *Mining* β / SE
Baby Boomer's	0.068***	-0.051***	0.160***	0.082***	-0.075***	-0.073***	-0.078***	-0.039
	(0.011)	(0.017)	(0.009)	(0.019)	(0.013)	(0.007)	(0.017)	(0.028)
Gen. X.	0.027**	-0.095***	0.187***	0.133***	-0.198***	-0.129***	-0.113***	-0.069**
	(0.012)	(0.020)	(0.010)	(0.021)	(0.014)	(0.009)	(0.017)	(0.035)
Millennial's	-0.195***	-0.340***	-0.060***	-0.070***	-0.475***	-0.362***	-0.331***	-0.023
	(0.015)	(0.033)	(0.013)	(0.027)	(0.023)	(0.016)	(0.022)	(0.048)
Constant	9.664***	10.307***	9.834***	10.095***	10.421***	10.296***	10.187***	10.571***
	(0.010)	(0.015)	(0.008)	(0.017)	(0.011)	(0.006)	(0.015)	(0.024)
Observations	65816	16774	121059	18700	29238	82557	26947	4443
R^2	0.006	0.007	0.007	0.004	0.019	0.007	0.008	0.001

Notes: This table presents estimates of equation (4.3), excluding covariates and fixed effects. The dependent variable is log wages. The specification estimated is: $\log w^i_{j,t} = \gamma^i_g + \varepsilon^i_{j,t}$. $\varepsilon^i_{j,t}$ are clustered by state and by year. The data from the CPS and BEA are merged at the state geographic level and 1 digit industry code as described in Appendix C.1. The generation variables are all dummy variables defined on the basis of date of birth, as per Table 4.1. The omitted category is the Silent Generation. Standard Errors are in parenthesis. *** Significant at the 1 percent level.** Significant at the 5 percent level. * Significant at the 10 percent level.

Retail, Services, and Finance. But, less in Wholesale, Utilities, Manufacturing, and Construction. Wages in mining remained about the same. With the exception of finance, where wages were 17% higher, the differences were mostly ±7%.[7] While, this consistency of absolute magnitude is perhaps of interest in its own right, here it suffices to note that these are large differences and imply substantial differences in lifetime consumption levels. Looking now at the second row we note that, again, Retail, Services, and Finance have positive coefficients and that all other sectors are associated with a reduction in the average wage. Moreover, the coefficients on average now seem to larger in magnitude with the average Gen X'er working in the Utilities sector earning around 18% less than their equivalent in the Silent Generation. Likewise, wages in Finance and Services are now 21% and 14% higher. When we look at the results for Millennials in the third row, this pattern breaks down. Now across all sectors the coefficient is negative and in some cases extremely large such as Utilities, where the implied reduction in earnings is 38%.

[7]Given the regressors are dummies, we can not interpret the changes directly in percentage terms, instead we calculate 17% = 100*(exp(0.160) − 1).

Table 4.3: Generational Differences in Real Wages by Industry: Fixed Effect Estimates

	(1) *Retail* β / SE	(2) *Wholesale* β / SE	(3) *Services* β / SE	(4) *Finance* β / SE	(5) *Utilities* β / SE	(6) *Manufacturing* β / SE	(7) *Construction* β / SE	(8) *Mining* β / SE
Baby Boomer's	-0.031	-0.133**	-0.011	-0.059	-0.127*	-0.177***	-0.159***	-0.146***
	(0.056)	(0.047)	(0.037)	(0.044)	(0.056)	(0.040)	(0.036)	(0.030)
Gen. X.	-0.185**	-0.302***	-0.130**	-0.170**	-0.314***	-0.367***	-0.313***	-0.347***
	(0.059)	(0.065)	(0.049)	(0.066)	(0.065)	(0.049)	(0.045)	(0.056)
Millennial's	-0.482***	-0.631***	-0.465***	-0.488***	-0.637***	-0.693***	-0.584***	-0.428***
	(0.057)	(0.047)	(0.053)	(0.056)	(0.082)	(0.054)	(0.055)	(0.059)
Constant	9.799***	10.409***	10.042***	10.274***	10.485***	10.408***	10.304***	10.709***
	(0.044)	(0.037)	(0.029)	(0.033)	(0.042)	(0.028)	(0.032)	(0.022)
Observations	65816	16774	121059	18700	29238	82557	26947	4442
R^2	0.030	0.038	0.046	0.059	0.033	0.052	0.048	0.077
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents estimates of equation (4.3), excluding covariates and including state and year fixed effects. The dependent variable is log wages. The specification estimated is: $\log w^i_{j,t} = \gamma^i_g + \delta^i_t + \delta^i_s + \varepsilon^i_{j,t}$. $\varepsilon^i_{j,t}$ are clustered by state and by year. The data from the CPS and BEA are merged at the state geographic level and 1 digit industry code as described in Appendix C.1. The generation variables are all dummy variables defined on the basis of date of birth, as per Table 4.1. The omitted category is the Silent Generation. Standard Errors are in parenthesis. *** Significant at the 1 percent level.** Significant at the 5 percent level. * Significant at the 10 percent level.

Of course, one explanation for these changes is that the average worker in sectors such as Finance has changed over time. Thus, increased earnings in Finance may reflect declining numbers of accounting clerks compared to investment bankers or financial advisers. Tables 4.3, 4.4, and 4.7 introduce fixed effects and our demographic controls such that the intergenerational comparison is now a more precise one. Looking at Table 4.3 now all of the coefficients are negative. Moreover, the absolute magnitude has increased such that the earnings of Boomers are 0–16% lower while those of Gen X. and Millenials are 18.5–31% and 35–50% lower respectively. Such differences may seem implausibly large, and when we include other controls in Table 4.5 the estimated changes are now lower, with the largest differences being for Millennials working in Manufacturing who make around 30% less than the Silents conditional on education, age, gender, and race.

When we include state and year fixed-effects as well as controlling for observables the estimated size and precision of our estimates decreases, perhaps because we are asking too much of the data. However, there is still a consistent pattern of lower

Table 4.4: Generational Differences in Real Wages by Industry: Including Covariates and Fixed Effects

	(1) *Retail* β / SE	(2) *Wholesale* β / SE	(3) *Services* β / SE	(4) *Finance* β / SE	(5) *Utilities* β / SE	(6) *Manufacturing* β / SE	(7) *Construction* β / SE	(8) *Mining* β / SE
Baby Boomer's	0.033	0.028	0.038	0.041	0.028	-0.002	-0.051	-0.008
	(0.030)	(0.037)	(0.027)	(0.037)	(0.033)	(0.010)	(0.029)	(0.044)
Gen. X.	0.017	0.046	0.037	0.036	0.009	-0.013	-0.047	-0.096*
	(0.049)	(0.050)	(0.052)	(0.063)	(0.051)	(0.024)	(0.044)	(0.043)
Millennial's	-0.038	0.007	-0.070	-0.047	-0.026	-0.102**	-0.118*	-0.002
	(0.073)	(0.072)	(0.079)	(0.082)	(0.079)	(0.033)	(0.052)	(0.062)
Age	0.091***	0.087***	0.075***	0.082***	0.099***	0.075***	0.065***	0.062***
	(0.003)	(0.006)	(0.005)	(0.005)	(0.006)	(0.002)	(0.005)	(0.008)
Age Sq	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
African American	-0.176***	-0.303***	-0.194***	-0.178***	-0.237***	-0.211***	-0.372***	-0.257*
	(0.025)	(0.037)	(0.021)	(0.025)	(0.020)	(0.011)	(0.037)	(0.115)
Hispanic	-0.131***	-0.308***	-0.180***	-0.162***	-0.173***	-0.267***	-0.248***	-0.093
	(0.022)	(0.043)	(0.023)	(0.020)	(0.018)	(0.023)	(0.034)	(0.058)
High School Graduate	0.284***	0.319***	0.452***	0.317***	0.246***	0.322***	0.300***	0.276***
	(0.035)	(0.027)	(0.023)	(0.040)	(0.030)	(0.016)	(0.019)	(0.035)
College	0.385***	0.411***	0.542***	0.464***	0.401***	0.521***	0.351***	0.400***
	(0.023)	(0.035)	(0.025)	(0.021)	(0.022)	(0.036)	(0.031)	(0.053)
Female	-0.589***	-0.488***	-0.450***	-0.366***	-0.378***	-0.479***	-0.436***	-0.440***
	(0.052)	(0.045)	(0.030)	(0.036)	(0.023)	(0.039)	(0.053)	(0.037)
Constant	7.851***	8.197***	8.067***	8.259***	8.032***	8.450***	8.617***	9.034***
	(0.129)	(0.136)	(0.184)	(0.181)	(0.179)	(0.071)	(0.119)	(0.145)
Observations	65816	16774	121059	18700	29238	82557	26947	4442
R^2	0.179	0.227	0.230	0.213	0.168	0.284	0.134	0.188
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents estimates of equation (4.3), including covariates but not fixed effects. The dependent variable is log wages. The specification estimated is: $\log w^i_{j,t} = \gamma^i_g + X'_{jt}\beta^i + \delta^i_t + \delta^i_s + \varepsilon^i_{j,t}$. $\varepsilon^i_{j,t}$ are clustered by state and by year. The data from the CPS and BEA are merged at the state geographic level and 1 digit industry code as described in Appendix C.1. The generation variables are all dummy variables defined on the basis of date of birth, as per Table 4.1. The omitted category is the Silent Generation. X contains education, race, gender, and age variables. Standard Errors are in parenthesis. *** Significant at the 1 percent level.** Significant at the 5 percent level. * Significant at the 10 percent level.

earnings amongst Millennials, particularly in the Manufacturing and Construction sectors. To understand the aggregate differences between generations we pool across industries and estimate equation (4.3) for the economy as a whole. The results are reported in columns 1–3 of Table 4.9 and 4.10. These result suggest that overall Boomers' wages have been higher than those of the Silents with Millennials having lower wages both conditionally and unconditionally than the three preceding generations. The results for Gen. X are not always significant but taken together suggest that, again allowing for both observables and unobservables, that their earnings were less than those of the Boomers or the Silents.

Taken together, our analysis of the individual level data suggest that even controlling for changes in the composition of the workforce and state and year fixed effects we still find a downwards trend in real median earnings from the Silent generation and the Boomers to the Millenials. We now consider why this might be the case.

4.3.3 Explanations

Table 4.6 reports the results of an analogous specification to equation (4.3) except that now we disaggregate by generation and pool over industries. This allows us to get a sense of the differences in the the effects of observables across generations. We begin by considering the effects of education and age (as our proxy for experience). Looking first at age we can see that the coefficients on Age and Age^2 are reasonably consistent across the first three generations and then starkly different for Millennials. Allowing for the differences in the constant term, this suggests that Millennial wages start lower, converge towards those of Gen X'ers before falling away after age 30.[8] Given we include year effects this different profile would be consistent with a disproportionate effect of the Financial Crisis on Millennials.

Perhaps most interesting is that the positive impact of education on wages first grew, as the literature on skills-biased technological change documents (Acemoglu and Autor, 2011), but has fallen for Millennials with the returns to college attendance lower for Millennials than for Boomer's. The return to high-school graduation is also lower than that for Gen. X'ers. It is important not to over-interpret a single specification given the substantial literature employing much richer specifications which document the increasing demand for college graduates and increasing returns to education (Acemoglu and Autor, 2011). Our interpretation of these results is

[8]See Figure C.5 in the Appendix.

Table 4.5: Generational Differences in the Labour share by Industry: Including Covariates

	(1) *Retail* β / SE	(2) *Wholesale* β / SE	(3) *Services* β / SE	(4) *Finance* β / SE	(5) *Utilities* β / SE	(6) *Manufacturing* β / SE	(7) *Construction* β / SE	(8) *Mining* β / SE
Baby Boomer's	0.003***	0.008***	0.072***	0.400***	0.317***	-0.035***	0.013***	0.018***
	(0.000)	(0.000)	(0.000)	(0.006)	(0.006)	(0.001)	(0.000)	(0.002)
Gen. X.	-0.012***	0.001***	0.025***	0.555***	0.421***	-0.145***	-0.021***	-0.036***
	(0.000)	(0.000)	(0.000)	(0.009)	(0.007)	(0.001)	(0.001)	(0.003)
Millennial's	-0.025***	-0.007***	-0.025***	0.478***	0.339***	-0.194***	-0.060***	-0.052***
	(0.000)	(0.000)	(0.000)	(0.011)	(0.009)	(0.001)	(0.001)	(0.003)
Age	0.002***	0.002***	0.013***	0.049***	0.030***	0.001***	0.007***	0.004***
	(0.000)	(0.000)	(0.000)	(0.002)	(0.002)	(0.000)	(0.000)	(0.001)
Age Sq	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
African American	-0.000	-0.001***	-0.001***	-0.010	0.038***	-0.008***	-0.002***	-0.004
	(0.000)	(0.000)	(0.000)	(0.007)	(0.008)	(0.001)	(0.001)	(0.003)
Hispanic	-0.001***	-0.001***	-0.003***	-0.035***	0.010	-0.017***	-0.005***	-0.015***
	(0.000)	(0.000)	(0.000)	(0.006)	(0.008)	(0.001)	(0.000)	(0.002)
High School Graduate	-0.003***	-0.000	-0.008***	-0.001	0.001	-0.040***	-0.007***	-0.021***
	(0.000)	(0.000)	(0.000)	(0.011)	(0.007)	(0.001)	(0.000)	(0.002)
College	0.000	0.000	0.000	0.019***	0.024***	-0.001**	-0.000	-0.002
	(0.000)	(0.000)	(0.000)	(0.005)	(0.007)	(0.000)	(0.000)	(0.002)
Female	-0.000	0.000	0.000**	-0.010**	0.008	-0.002***	-0.000	0.004*
	(0.000)	(0.000)	(0.000)	(0.005)	(0.006)	(0.000)	(0.000)	(0.002)
Constant	0.026***	-0.028***	-0.096***	-1.171***	-0.829***	0.283***	0.006**	0.047***
	(0.001)	(0.001)	(0.002)	(0.039)	(0.038)	(0.003)	(0.002)	(0.012)
Observations	60459	14623	124184	13265	20877	74404	27996	3154
R^2	0.598	0.507	0.645	0.237	0.078	0.607	0.592	0.403

Notes: This table presents estimates of equation (4.3), including covariates but not fixed effects. The dependent variable is the labour share of value added. The specification estimated is: $\lambda_g^{s,t} = \gamma_g^i + X_{jt}'\beta^i + \varepsilon_{j,t}^i$. $\varepsilon_{j,t}^i$ are clustered by state and by year. The data from the CPS and BEA are merged at the state geographic level and 1 digit industry code as described in Appendix C.1. The generation variables are all dummy variables defined on the basis of date of birth, as per Table 4.1. The omitted category is the Silent Generation. Standard Errors are in parenthesis. *** Significant at the 1 percent level.** Significant at the 5 percent level. * Significant at the 10 percent level.

that, very broadly, while the college wage premium has increased across previous generations that it now seems to be falling for Millennials, perhaps limiting wage growth despite growing college attendance rates.

Thus, it is hard to explain the intergenerational decline in median earnings with an appeal to the changing returns to human capital. It does not seem to be the case that later generations have earned much larger returns to education. The other key change in the labour market in the period has been increased participation by women. This is born out by the reduction in the magnitude of the Female coefficient which is just over one quarter for Millennials down from 55% for Silents. That is, there continues to be a substantial gender pay gap between Millennials, but it is substantially smaller than that of previous generations. This is as expected given

Table 4.6: Regression on Log Wage by Generation

	(1) *Silent* β / SE	(2) *Boomer's* β / SE	(3) *Gen. X* β / SE	(4) *Millenials* β / SE	(5) *Silent* β / SE	(6) *Boomer's* β / SE	(7) *Gen. X* β / SE	(8) *Millenials* β / SE
Age	0.089***	0.075***	0.129***	0.486***	0.083***	0.068***	0.122***	0.454***
	(0.010)	(0.005)	(0.021)	(0.038)	(0.009)	(0.005)	(0.020)	(0.031)
Age Sq	-0.001***	-0.001***	-0.002***	-0.008***	-0.001***	-0.001***	-0.001***	-0.007***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
African American	-0.220***	-0.222***	-0.199***	-0.189**	-0.215***	-0.227***	-0.200***	-0.185**
	(0.039)	(0.016)	(0.021)	(0.024)	(0.037)	(0.016)	(0.021)	(0.023)
Hispanic	-0.260***	-0.220***	-0.174***	-0.086**	-0.264***	-0.223***	-0.177***	-0.093**
	(0.039)	(0.016)	(0.021)	(0.010)	(0.037)	(0.014)	(0.021)	(0.011)
High School Graduate	0.315***	0.319***	0.392***	0.365***	0.324***	0.320***	0.384***	0.356***
	(0.012)	(0.035)	(0.020)	(0.005)	(0.012)	(0.031)	(0.020)	(0.011)
College	0.385***	0.488***	0.540***	0.475**	0.416***	0.492***	0.525***	0.457**
	(0.031)	(0.029)	(0.035)	(0.050)	(0.025)	(0.024)	(0.032)	(0.050)
Female	-0.785***	-0.539***	-0.453***	-0.310***	-0.685***	-0.480***	-0.416***	-0.257***
	(0.069)	(0.022)	(0.025)	(0.024)	(0.067)	(0.019)	(0.026)	(0.018)
Retail Trade					0.000	0.000	0.000	0.000
					(.)	(.)	(.)	(.)
Wholesale Trade					0.377***	0.369***	0.343***	0.344**
					(0.033)	(0.023)	(0.014)	(0.035)
Services					0.132**	0.169***	0.189***	0.164***
					(0.042)	(0.021)	(0.016)	(0.013)
Finance					0.346***	0.382***	0.386***	0.383***
					(0.039)	(0.017)	(0.021)	(0.018)
Utilities & Transportation					0.487***	0.437***	0.338***	0.336**
					(0.015)	(0.035)	(0.018)	(0.039)
Manufacturing					0.444***	0.394***	0.343***	0.324***
					(0.037)	(0.021)	(0.014)	(0.026)
Construction					0.222***	0.211***	0.248***	0.299***
					(0.048)	(0.019)	(0.016)	(0.011)
Mining					0.606***	0.658***	0.609***	0.877***
					(0.063)	(0.047)	(0.053)	(0.034)
Constant	8.252***	8.340***	7.226***	1.899*	8.060***	8.228***	7.157***	2.200**
	(0.269)	(0.100)	(0.377)	(0.557)	(0.261)	(0.109)	(0.360)	(0.462)
Observations	66183	177964	94783	26604	66183	177964	94783	26604
R^2	0.250	0.214	0.235	0.174	0.281	0.241	0.254	0.199
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

*Notes:*This table presents estimates of equation (4.3) but pooling across industries and disaggregating instead by industry, including covariates and fixed effects. The dependent variable is log wages. The specification estimated is: $\log w^g_{j,t} = X'_{jt}\beta^g + \delta_t + \delta_s + \varepsilon^g_{j,t}$. $\varepsilon^g_{j,t}$ are clustered by state and by year. Where $c^i_{j,t}$, the dependent variable, is the log wage in columns (1) to (3) and the labour share of value added in columns (4) to (6). The data from the CPS and BEA are merged at the state geographic level and 1 digit industry code as described in Appendix C.1. The generation variables are all dummy variables defined on the basis of date of birth, as per Table 4.1. The omitted category is the Silent Generation. Standard Errors are in parenthesis. *** Significant at the 1 percent level.** Significant at the 5 percent level. * Significant at the 10 percent level.

the findings of the large literature on the US gender pay gap (Goldin, 2014). But, as an explanation for the intergenerational decline in median wages it is most useful to consider it in light of the analysis of Guvenen et al. (2017). Recall, that their analysis shows that while women's share of cohort lifetime earnings nearly doubles across the 27 cohorts they study that much of this increase is due to increased labour force attachment and that median women's earnings have grown slowly and

Table 4.7: Generational Differences in Real Wages by Industry: Including Covariates

	(1) *Retail* β / SE	(2) *Wholesale* β / SE	(3) *Services* β / SE	(4) *Finance* β / SE	(5) *Utilities* β / SE	(6) *Manufacturing* β / SE	(7) *Construction* β / SE	(8) *Mining* β / SE
Baby Boomer's	0.001	-0.072***	0.052***	0.046**	-0.120***	-0.115***	-0.113***	-0.051*
	(0.012)	(0.018)	(0.009)	(0.020)	(0.014)	(0.007)	(0.018)	(0.029)
Gen. X.	-0.045***	-0.124***	0.067***	0.045*	-0.259***	-0.220***	-0.127***	-0.156***
	(0.015)	(0.024)	(0.011)	(0.025)	(0.017)	(0.009)	(0.021)	(0.040)
Millennial's	-0.112***	-0.199***	-0.015	-0.008	-0.370***	-0.357***	-0.207***	-0.006
	(0.019)	(0.037)	(0.014)	(0.032)	(0.026)	(0.017)	(0.027)	(0.054)
Age	0.090***	0.082***	0.077***	0.084***	0.090***	0.068***	0.063***	0.053***
	(0.003)	(0.005)	(0.002)	(0.004)	(0.004)	(0.002)	(0.004)	(0.008)
Age Sq	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***	-0.001***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
African American	-0.148***	-0.273***	-0.151***	-0.125***	-0.215***	-0.225***	-0.372***	-0.254***
	(0.013)	(0.024)	(0.008)	(0.020)	(0.015)	(0.009)	(0.024)	(0.064)
Hispanic	-0.101***	-0.292***	-0.140***	-0.108***	-0.170***	-0.252***	-0.218***	-0.061*
	(0.010)	(0.020)	(0.008)	(0.018)	(0.013)	(0.008)	(0.013)	(0.034)
High School Graduate	0.238***	0.272***	0.419***	0.261***	0.169***	0.267***	0.269***	0.272***
	(0.009)	(0.015)	(0.008)	(0.019)	(0.012)	(0.006)	(0.012)	(0.026)
College	0.393***	0.418***	0.561***	0.477***	0.400***	0.537***	0.369***	0.398***
	(0.012)	(0.017)	(0.006)	(0.015)	(0.012)	(0.007)	(0.020)	(0.031)
Female	-0.595***	-0.487***	-0.461***	-0.386***	-0.374***	-0.487***	-0.426***	-0.430***
	(0.007)	(0.013)	(0.005)	(0.013)	(0.010)	(0.006)	(0.020)	(0.032)
Constant	7.969***	8.511***	8.017***	8.260***	8.584***	8.828***	8.747***	9.276***
	(0.052)	(0.096)	(0.041)	(0.090)	(0.075)	(0.041)	(0.080)	(0.167)
Observations	65816	16774	121059	18700	29238	82557	26947	4443
R^2	0.166	0.208	0.217	0.192	0.152	0.263	0.106	0.138

Notes: This table presents estimates of equation (4.3), including covariates but not fixed effects. The dependent variable is log wages. The specification estimated is: $\log w^i_{j,t} = \gamma^i_g + X'_{jt}\beta^i + \varepsilon^i_{j,t}$. $\varepsilon^i_{j,t}$ are clustered by state and by year. The data from the CPS and BEA are merged at the state geographic level and 1 digit industry code as described in Appendix C.1. The generation variables are all dummy variables defined on the basis of date of birth, as per Table 4.1. The omitted category is the Silent Generation. X contains education, race, gender, and age variables. Standard Errors are in parenthesis. *** Significant at the 1 percent level.** Significant at the 5 percent level. * Significant at the 10 percent level.

remain at a substantially lower level than those of men. Similarly, there is limited evidence of substantial intergenerational changes in racial earnings inequality. The small (and statistically insignificant) decline in the coefficient on the African American dummy variable suggests that there has been little improvement in that earnings gap between the Silent Generation and Millennials.[9] On the other hand, the gap for Hispanic Americans has reduced markedly – from around 23% for Silents to just over 8% for Millennials.

It seems then that notwithstanding the decline in the Hispanic earnings gap and increased female labour force participation, that the correlates of earnings have been

[9]Lang and Manove (2011) show that the racial earnings gap is increased when a richer set of educational controls are included. Fryer (2011) shows that there is good reason to believe that it has declined for those born more recently.

broadly consistent from one generation to the next. Thus, understanding the decline in median wages requires another explanation. To this end, we now consider the role of the labour share.

Table 4.8: Generational Differences in the Labour share by Industry: Unconditional Estimates

	(1) ***Retail*** β / SE	(2) ***Wholesale*** β / SE	(3) ***Services*** β / SE	(4) ***Finance*** β / SE	(5) ***Utilities*** β / SE	(6) ***Manufacturing*** β / SE	(7) ***Construction*** β / SE	(8) ***Mining*** β / SE
Baby Boomer's	0.010***	0.009***	0.089***	0.376***	0.277***	0.006***	0.028***	0.038***
	(0.000)	(0.000)	(0.000)	(0.005)	(0.004)	(0.001)	(0.001)	(0.002)
Gen. X.	0.003***	0.003***	0.052***	0.406***	0.286***	-0.082***	0.005***	-0.005***
	(0.000)	(0.000)	(0.000)	(0.005)	(0.004)	(0.001)	(0.001)	(0.002)
Millennial's	-0.008***	-0.008***	-0.011***	0.203***	0.109***	-0.110***	-0.034***	-0.016***
	(0.000)	(0.000)	(0.000)	(0.005)	(0.002)	(0.001)	(0.001)	(0.002)
Constant	0.036***	0.018***	0.105***	0.082***	0.043***	0.131***	0.081***	0.046***
	(0.000)	(0.000)	(0.000)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
Observations	60459	14623	124184	13265	20877	74404	27996	3154
R^2	0.197	0.338	0.467	0.154	0.051	0.265	0.282	0.225

Notes: This table presents estimates of equation (4.3), excluding covariates and fixed effects. The dependent variable isthe labour share of value added. The specification estimated is: $\lambda_g^{s,t} = \gamma_g^i + \varepsilon_{j,t}^i$. $\varepsilon_{j,t}^i$ are clustered by state and by year. The data from the CPS and BEA are merged at the state geographic level and 1 digit industry code as described in Appendix C.1. The generation variables are all dummy variables defined on the basis of date of birth, as per Table 4.1. The omitted category is the Silent Generation. Standard Errors are in parenthesis. *** Significant at the 1 percent level.** Significant at the 5 percent level. * Significant at the 10 percent level.

Table 4.8 reports estimates of equation (4.3) omitting the fixed-effects and controls. We can see that the Boomer's labour share was consistently higher than that of the Silents. Gen. X'ers had lower labour shares than Boomers, with smaller increases relative to the Silents in sectors such as Retail, Wholesale, or Construction and reductions in Manufacturing and Mining. Like with wages, the Finance and Utilities sectors are the exceptions showing increases in the labour share. Millennials have lower labour shares than all previous generations in every sector but Finance and Utilities, where they have lower labour shares than Boomers and Gen. X'ers but higher shares than Silents. These results are consistent with and exhibit similar patterns to the results for wages above, suggesting that intergenerational declines in median wages may reflect declines in the labour share.

Controlling for observables has a limited impact on the results, as can be seen in Table 4.5. The estimated γ coefficients are now larger, with a similar pattern across industries as in Table 4.8. Baby Boomer's still have, except in Manufacturing, the highest labour share of any generation, the ordering of the Gen. X'ers and the Silents is evenly split across industries. Silents have higher labour shares in Manufacturing, Construction, Mining, and Retail while Gen. X'ers do better in the other industries. Millennials continue to have a much lower labour share in every industry except Finance and Utilities, although in these industries, like in all others, the Millennial Labour share is lower than that of the Boomers and Gen. X'ers.

Looking at the coefficients on the control variables we see some interesting patterns. Firstly, individuals' labour shares follow the familiar parabolic path over the life-cycle. African-Americans and Hispanics both receive a consistently lower labour share consistent with their lower earnings conditional on education and experience. High School Graduates also have a lower share perhaps reflecting the greater substitutability of low-skilled labour with machinery as argued by Autor and Salomons (2018). That college attendees working in the Finance or Utilities sectors have higher labour shares in all generations, as well as these sectors not showing the same decline in labour shares may similarly reflect the the nature of technical change in these industries.

To close we consider results pooling across industries to analyse overall differences between generations. The results are reported in Columns 4–6 of Tables 4.9 and 4.10. Here we see, again, that the Baby Boomer's had the largest labour share followed by the Gen. X'ers. Without fixed effects, our preferred specification given how we construct the labour share data, the Millennials have then a lower labour share than the Silents who have a lower share than the Boomers and the Gen. X'ers. When we include them, the Silents and the Millennials have no significant difference in Labour share, suggesting that the substantial growth in

Table 4.9: Generational Differences in Wage and the Labour Share: Pooled Estimates

	(1) Wage β / SE	(2) Wage β / SE	(3) Wage β / SE	(4) Labour Share β / SE	(5) Labour Share β / SE	(6) Labour Share β / SE
Baby Boomer's	0.000	-0.058***	-0.039***	0.068***	0.056***	0.043***
	(0.004)	(0.004)	(0.004)	(0.000)	(0.001)	(0.000)
Gen. X.	-0.048***	-0.124***	-0.079***	0.042***	0.024***	0.004***
	(0.005)	(0.006)	(0.006)	(0.001)	(0.001)	(0.001)
Millennial's	-0.314***	-0.224***	-0.157***	-0.017***	-0.020***	-0.040***
	(0.007)	(0.008)	(0.008)	(0.001)	(0.001)	(0.001)
Age		0.084***	0.078***		0.012***	0.010***
		(0.001)	(0.001)		(0.000)	(0.000)
Age Sq		-0.001***	-0.001***		-0.000***	-0.000***
		(0.000)	(0.000)		(0.000)	(0.000)
African American		-0.189***	-0.191***		0.011***	-0.000
		(0.005)	(0.005)		(0.001)	(0.001)
Hispanic		-0.182***	-0.176***		-0.006***	-0.010***
		(0.004)	(0.004)		(0.001)	(0.001)
High School Graduate		0.281***	0.286***		-0.006***	-0.026***
		(0.004)	(0.004)		(0.001)	(0.000)
College		0.513***	0.511***		0.026***	0.003***
		(0.004)	(0.004)		(0.001)	(0.001)
Female		-0.545***	-0.488***		0.017***	-0.000
		(0.003)	(0.003)		(0.000)	(0.000)
Retail Trade			0.000			0.000
			(.)			(.)
Wholesale Trade			0.374***			-0.024***
			(0.007)			(0.000)
Services			0.185***			0.119***
			(0.004)			(0.000)
Finance			0.396***			0.384***
			(0.007)			(0.002)
Utilities & Transportation			0.432***			0.247***
			(0.006)			(0.003)
Manufacturing			0.399***			0.061***
			(0.004)			(0.000)
Construction			0.235***			0.044***
			(0.006)			(0.000)
Mining			0.624***			0.018***
			(0.011)			(0.001)
Constant	10.071***	8.324***	8.115***	0.091***	-0.122***	-0.127***
	(0.004)	(0.022)	(0.022)	(0.000)	(0.003)	(0.003)
Observations	365534	365534	365534	338962	338962	338962
R^2	0.007	0.209	0.234	0.040	0.062	0.391

Notes: This table presents estimates of equation (4.3) but pooling across industries, including covariates but not fixed effects. The specification estimated is: $c^i_{j,t} = \gamma^i_g + X'_{jt}\beta^i + \varepsilon^i_{j,t}$. $\varepsilon^i_{j,t}$ are clustered by state and by year. Where $c^i_{j,t}$, the dependent variable, is the log wage in columns (1) to (3) and the labour share of value added in columns (4) to (6). The data from the CPS and BEA are merged at the state geographic level and 1 digit industry code as described in Appendix C.1. The generation variables are all dummy variables defined on the basis of date of birth, as per Table 4.1. The omitted category is the Silent Generation. Standard Errors are in parenthesis. *** Significant at the 1 percent level.** Significant at the 5 percent level. * Significant at the 10 percent level.

the labour share of those employed in the Finance and Utilities sectors is sufficient to offset the decrease in all other sectors. It is useful to think about the exceptionalness of these sectors, in light of the findings of Guvenen et al. (2017) who show that only the top decile has seen wage growth across cohorts. Consistent with both

Table 4.10: Regression on Wage and Labour Share

	(1) Wage β / SE	(2) Wage β / SE	(3) Wage β / SE	(4) Labour Share β / SE	(5) Labour Share β / SE	(6) Labour Share β / SE
Baby Boomer's	-0.106*	0.019	0.013	0.072***	0.065***	0.063***
	(0.048)	(0.020)	(0.018)	(0.020)	(0.013)	(0.013)
Gen. X.	-0.276***	0.013	0.011	0.051*	0.045**	0.043**
	(0.057)	(0.036)	(0.034)	(0.022)	(0.019)	(0.018)
Millennial's	-0.624***	-0.064	-0.059	-0.006	0.009	0.011
	(0.055)	(0.053)	(0.050)	(0.027)	(0.023)	(0.022)
Age		0.087***	0.080***		0.012***	0.011***
		(0.003)	(0.003)		(0.001)	(0.001)
Age Sq		-0.001***	-0.001***		-0.000***	-0.000***
		(0.000)	(0.000)		(0.000)	(0.000)
African American		-0.212***	-0.214***		0.012***	0.002
		(0.018)	(0.018)		(0.003)	(0.001)
Hispanic		-0.201***	-0.203***		-0.003	-0.002
		(0.016)	(0.016)		(0.003)	(0.001)
High School Graduate		0.336***	0.336***		0.011*	-0.005***
		(0.026)	(0.023)		(0.005)	(0.001)
College		0.500***	0.498***		0.023***	0.003**
		(0.028)	(0.023)		(0.002)	(0.001)
Female		-0.542***	-0.483***		0.018***	0.000
		(0.042)	(0.035)		(0.003)	(0.001)
Retail Trade			0.000			0.000
			(.)			(.)
Wholesale Trade			0.372***			-0.025***
			(0.017)			(0.003)
Services			0.173***			0.111***
			(0.018)			(0.006)
Finance			0.381***			0.385***
			(0.014)			(0.084)
Utilities & Transportation			0.423***			0.249**
			(0.034)			(0.077)
Manufacturing			0.400***			0.048*
			(0.020)			(0.021)
Construction			0.233***			0.039***
			(0.016)			(0.006)
Mining			0.660***			0.027**
			(0.042)			(0.009)
Constant	10.205***	8.065***	7.946***	0.085***	-0.174***	-0.209***
	(0.037)	(0.113)	(0.114)	(0.016)	(0.030)	(0.025)
Observations	365534	365534	365534	338962	338962	338962
R^2	0.033	0.222	0.247	0.088	0.110	0.441
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
State FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table presents estimates of equation (4.3) but pooling across industries, including covariates and fixed effects. The specification estimated is: $c^i_{j,t} = \gamma^i_g + X'_{jt}\beta^i + \delta_t + \delta_s + \varepsilon^i_{j,t}$. $\varepsilon^i_{j,t}$ are clustered by state and by year. Where $c^i_{j,t}$, the dependent variable, is the log wage in columns (1) to (3) and the labour share of value added in columns (4) to (6). The data from the CPS and BEA are merged at the state geographic level and 1 digit industry code as described in Appendix C.1. The generation variables are all dummy variables defined on the basis of date of birth, as per Table 4.1. The omitted category is the Silent Generation. Standard Errors are in parenthesis. *** Significant at the 1 percent level.** Significant at the 5 percent level. * Significant at the 10 percent level.

of these findings is the possibility is that the growth in wages and the labour share in these sectors is both driven by the highest earning in those sectors, and also is partly driving the growth in the incomes of the richest.

4.4 Variance Decomposition

Having found evidence for inter-generational differences in wages and the labour share we now ask how important these are relative to other changes in the US economy over the period we study. As well as the key change of increased female labour force participation and skill-biased technological change there have also been changes in the geographic distribution of economic activity away from the Rust-Belt and towards the South and West. There has also been substantial changes in the relative importance of different sectors, with the decline of manufacturing being one prominent example. In this section we employ a variance-decomposition analysis to understand the quantitative importance of these different trends, and specifically the relative importance of inter-generational differences.

Rather than focus on the individual coefficients γ and β we now ask instead how much of the variation in log wages or the labour share can be explained by each set of variables. If a particular variable or group of variables can explain a lot of the variation then it clearly an economically significant determinant. On the other hand, a precisely estimated coefficient that cannot explain much of the variation is not.

4.4.1 Variance Decomposition Estimators

To begin, following the approach of Gibbons et al. (2014), consider a simplified version of equation (4.3) in which we pool across industries:

$$c_{j,t} = d_s'\delta + X_{jt}'\beta + \varepsilon_{j,t} \tag{4.4}$$

We now write δ_s as $d_s'\delta$ as it will be useful to distinguish between the data and the parameters below. Here, we are abstracting for now from generation and industry, and focussing on the quantitative importance of where an individual lives and their characteristics. That is, we wish to distinguish between the idea that individuals in richer states earn more because the best and the brightest move to those states versus the hypothesis that it is the nature of the richer states themselves, whether it be natural resources, geography, or infrastructure, etc., that means that otherwise identical workers earn more there.

The complication emergences since it is both reasonable to believe that the state effects, δ_s, are a composite of the exogenous features of each state and the characteristics of those that live there. At the same time as maintaining that that the individual characteristics, such as education, will depend in part on where the individual lives as well as on their innate characteristics. A further complication is sorting: that people may endogenously relocate based on their characteristics.

Gibbons et al. (2014) discuss how alternative assumptions about what determines the state and individual effects give rise to a number of alternative variance-decomposition estimators. Their starting point is the *Raw Variance Share* (RVS), which estimates the effects of states as simply the R^2 of a simple regression which includes only state dummies. Thus, the RVS provides an upper-bound on what proportion of the variation of earnings is due to location, since it takes into account no other differences between workers than their location.

It is more plausible to assume that some of the variation in wages between areas is due to the composition of the labour force. An extreme approach is to focus only on the area effects that are uncorrelated with the individual effects, that is to assume that state characteristics are responsible for none of the variation in individual characteristics. As Gibbons et al. (2014) term this estimator the *Uncorrel-*

ated Variance Share (UVS) and it is most easily thought of as the difference in the R^2 of a regression including individual and area effects compared to a regression only including individual effects. Formally,

$$UVS = R^2(c_{j,t}; X'_{j,t}\hat{\beta}, d'_s\hat{\delta}) - R^2(c_{j,t}; X'_{j,t}\hat{\beta}) \tag{4.5}$$

Where $R^2(c_{j,t}; X'_{j,t}\hat{\beta}, d'_s\hat{\delta})$ is the R^2 from estimating equation (4.4), and and $R^2(c_{j,t}; X'_{j,t}\hat{\beta})$ is the R^2 is from a similar regression which does not include the state dummy variables.

An intermediate approach is rather than assuming that all of the variation is due to state characteristics (as for the RVS) or that state effects are only what cannot be explained by the (observed) characteristics of their populations (the UVS), is that there is a relationship between state characteristics and their populations. An intuitive approach is the *Correlated Variance Share* (CVS), which is the estimated variance share of location conditional on individual characteristics. That is, it's the variance share allowing for state and individual characteristics to be correlated. Substantively, this means it attributes to the area effect the effects due to the composition in terms of individual characteristics of the area, but not the effects of the area on its composition. Put differently, it attributes to the area the effects of the overall distribution of characteristics but does not allow for the possibility that the area may affect education outcomes, or induce sorting across areas. It may be computed as follows:

$$CVS = \frac{var(d'_s\hat{\delta})}{var(c_{j,t})} \tag{4.6}$$

A prominent related approach is the estimator of Abowd et al. (1999) which Gibbons et al. (2014) term the *Balanced Variance Share* (BVS). This last method builds upon CVS, but explicitly addresses the issue of sorting, or other ways in which areas may influence individual characteristics. Specifically, it includes the co-variance between location and individual characteristics. This, in effect, attributes half of the effects of sorting to the area effects (and the other half to individuals).

$$BVS = \frac{var(d_s'\hat{\delta}) + cov(d_s'\hat{\delta}, X_{j,t}'\hat{\beta})}{var(c_{j,t})} \tag{4.7}$$

We will present all four estimators since the RVS and the UVS capture the upper and lower limits of the area effects while comparing the CVS and the BVS to the UVS allows us to recover the correlation between individuals and areas, and the extent of endogenous sorting across areas.[10]

4.4.2 Variance Decomposition Estimates

We now return to our main model which is a variant of equation (4.3) which pools across industries but potentially includes industry fixed-effects. That is,

$$c_{j,t} = \gamma_g + X_{jt}'\beta + d_t'\delta_t + d_s'\delta_s + d_i'\delta_i + \varepsilon_{j,t} \tag{4.8}$$

where $d_i'\delta_i$ are industry fixed effects and all other terms are as in (4.3) and (4.4). Now, given that there are more than two sets of coefficients the formulae for BVS is slightly more complicated.[11] For an arbitrary variable $d_z'\delta_z$ we have that:

$$BVS(\hat{\delta}_z) = \frac{var(d_z'\hat{\delta}_z)}{var(c_{j,t})} + \sum_{z \neq w} \frac{cov(d_z'\hat{\delta}_z, d_w'\hat{\delta}_w)}{var(c_{j,t})} \tag{4.9}$$

[10]Note that, unlike the RVS, UVS, and CVS which all take values between 0 and 1, the BVS can take any value on the real line depending on the covariance terms.

[11]The expressions for CVS, RVS, and UVS are essentially as before. See Gibbons et al. (2014) for details.

Table 4.11 reports *UVS*, *CVS*, and *BVS* computed from estimates of (4.8). The first column reports results for a restricted model only including $d_t'\delta_t$ and $d_s'\delta_s$ which corresponds to the *RVS* estimate measure. Considering first the estimates for log wages in panel (a) we can see that location alone explains relatively little of the variation in wages across state – 1.8%. Note that all of the results in column (1) are identical since in the case of only one set of regressors all three estimators we consider collapse to the *RVS* (i.e. the R^2 of the state fixed effects). Column (2) reports results containing both individual characteristics and state and year fixed effects. Now, the explanatory power of the state fixed effects is lower, in every case less than 1%. The explanatory power of the individual characteristics is around 20%, with little variation across the different estimates. This figure seems intuitively plausible since we expect education, experience (age), and gender to be important predictors of earnings but there will remain considerable unobserved heterogeneity.

Column 3 additionally includes industry fixed effects. These themselves explain between 2.8–3.6% of the variance in earnings. There is now also more dispersion between the *BVS*, *CVS*, and *UVS* with *BVS-CVS* $\approx$ *CVS-UVS* $\approx$ 2.5% suggesting that both sorting across states and industries (*BVS-CVS)*and the state and industry composition of the labour market (*CVS-UVS*) are important. Interestingly, these effects are of a similar order to the direct variance share of states and industries suggesting that the indirect effects of states and industries on earnings due to differences in who works in them and sorting is similar to the direct effects of the characteristics of the states/industries themselves.

Having seen that state and industry are comparatively unimportant for wages compared to individual characteristics, Panel (a) of Table 4.12 reports estimates of equation (4.8) where the set of individual characteristics is disaggregated into edu-

Table 4.11: Variance Decomposition

	Year Dummies Only (1)	Plus Individual Characteristics (2)	Plus Industry Dummies (3)
Panel (a): Log Wage			
State Variance Share			
UVS	1.83	0.79	0.80
CVS	1.83	0.83	0.85
BVS	1.83	0.78	0.78
Individual Variance Share			
UVS		20.31	18.06
CVS		22.19	20.31
BVS		22.14	22.90
Industry Variance Share			
UVS			2.78
CVS			2.87
BVS			3.60
Panel (b): Labour Share			
State Variance Share			
UVS	5.03	1.15	1.11
CVS	5.03	1.22	1.18
BVS	5.03	1.26	1.26
Individual Variance Share			
UVS		5.93	4.40
CVS		6.35	4.50
BVS		6.39	6.64
Industry Variance Share			
UVS			32.85
CVS			41.78
BVS			42.13

Notes: This table reports the results of a variance decomposition based on a estimates of the following regression:

$$c_{j,t} = \gamma_g + X'_{jt}\beta + d'_t\delta_t + d'_s\delta_s + d'_i\delta_i + \varepsilon_{j,t} \tag{4.10}$$

The included individual characteristics are age, gender, race and education dummies in addition to the generation dummies. See Section 4.4.1 for definitions of *UVS*,*CVS*, and *BVS*. . All of the above numbers are percentages.

cation, generation, and demographic groups. Given the evidence of stagnating and indeed declining wages, we expect that the variance share of generation should be either 0 or weakly negative. First, looking down both columns we see that the inclusion of industry effects makes little difference. The impact of the education attainment variables ranges from 7.28-8.00% for the *UVS* to nearly 10.69-10.84% for the *BVS*. This suggests that education accounts for about half of the effect of the individual characteristics. Looking at the Demographic controls we can see estimates range from as much as 16.9-18.8% (*UVS*) to 10.1-10.9% (*BVS*). Crucially, the estimated share of generational differences is around 0, except for the *BVS* where it is -2.3 and -3.1% excluding and including industry dummies respectively. This suggests that the sum of the correlations between the generational dummies and education and demographic dummies is negative and that as such education, race, age, and gender explain less of the variation in wages for generations subsequent to the Silents.[12] Since the Total Sum of Squares for equation (4.8) must always be 1 for a given a set of estimates $\hat{\delta}$ the negative correlation with the observed demographic characteristics must be offset by a positive correlation elsewhere, potentially with an interaction of observed characteristics or unobserved characteristics.

Thus, the results for log wages suggest that differences between generations have limited explanatory power which is consistent with stagnating average real wages. We also found limited importance for state of residence and industry. The results for the labour share in panel (b) of Tables 4.11 and 4.12 point in the opposite direction. Looking at Table 4.11 we can see that state is more important than it was for wages, but the big change is that industry explains between one third and two fifths of the variance in the labour share. This compares with only around 5% for individual characteristics. When, as reported in Table 4.12 we separate out these individual characteristics we can see that now generation explains between 2.4 *UVS* and 5.4% *BVS* without industry dummies, and between 2.3 and 0.8% with them. The comparatively small effect for the *BVS* with industry dummies suggests that there is a negative correlation between the explanatory power of industry and generation.

[12]To see this, note that here we can write equation (4.9) as:

$$BVS(\hat{\delta}_{\text{gen}}) = 0 + cov(d'_{\text{gen}}\hat{\delta}_{\text{gen}}, d'_{\text{educ}}\hat{\delta}_{\text{educ}}) + cov(d'_{\text{gen}}\hat{\delta}_{\text{gen}}, d'_{\text{race}}\hat{\delta}_{\text{race}}) + cov(d'_{\text{gen}}\hat{\delta}_{\text{gen}}, d'_{\text{female}}\hat{\delta}_{\text{female}}) + cov(d'_{\text{gen}}\hat{\delta}_{\text{gen}}, d'_{\text{age}}\hat{\delta}_{\text{age}})$$

Given that all of δ_{gen} are binary variables it must be the case that the variance explained by the other characteristics is lower when one element of δ_{gen} is 1.

Table 4.12: Variance Decomposition: Individual Breakdown

	Plus Individual Characteristics (1)	Plus Industry Dummies (2)
Panel (a): Log Wage		
Education Variance Share		
UVS	8.00	7.28
CVS	10.17	9.94
BVS	10.84	10.69
Generation Variance Share		
UVS	0.02	0.01
CVS	0.02	0.02
BVS	-2.29	-3.11
Demographic Controls Variance Share		
UVS	18.76	16.88
CVS	10.17	9.94
BVS	10.12	10.86
Panel (b): Labour Share		
Education Variance Share		
UVS	0.50	0.02
CVS	0.61	0.02
BVS	0.70	0.57
Generation Variance Share		
UVS	2.42	2.25
CVS	2.98	2.77
BVS	5.38	0.77
Demographic Controls Variance Share		
UVS	2.24	1.13
CVS	0.61	0.02
BVS	0.66	0.72

Notes: This table reports the results of a variance decomposition based on a estimates of the following regression:

$$c_{j,t} = \gamma_g + X'_{jt}\beta + d'_t\delta_t + d'_s\delta_s + d'_i\delta_i + \varepsilon_{j,t} \quad (4.11)$$

The included individual characteristics, in X are age, gender, race and education dummies in addition to the generation dummies. See Section 4.4.1 for definitions of UVS,CVS, and BVS. . All of the above numbers are percentages. We consider *standard controls* to be gender, race, age and education variables. *Generation* includes the generation dummies and *education* is comprised of the dummies relating to education outcome. All of the above numbers are percentages.

To better understand the changes in the what determines earnings across generations without too much complication Table 4.13 reports decomposition estimates by generation. Looking first at the results for wages, we can see that whilst always relatively unimportant that the impact of state on earnings has fallen from around 1.1% for the Silents to around 0.6% for Millenials. Indeed this pattern of declining explanatory power also appears for the individual and industry variance shares for which the CVS falls from 19.6 to 16.4% and 3.4 to 3% respectively. For the labour share, the key change is that the role of industry almost doubles from around 35%

Table 4.13: Variance Decomposition: For each Generation

	Silent	Boomer's	Gen. X	Millennial
Panel (a): Log Wage				
State Variance Share				
UVS	1.05	0.89	0.77	0.55
CVS	1.10	0.94	0.84	0.60
BVS	1.13	0.88	0.75	0.66
Individual Variance Share				
UVS	18.43	16.68	16.83	13.83
CVS	19.60	20.87	20.84	16.44
BVS	20.55	21.42	21.52	17.61
Industry Variance Share				
UVS	3.31	2.89	2.28	2.90
CVS	3.37	2.99	2.37	2.98
BVS	4.29	3.56	3.10	4.05
Panel (b): Labour Share				
State Variance Share				
UVS	0.98	1.35	1.51	1.21
CVS	1.03	1.43	1.59	1.28
BVS	1.45	1.50	1.58	1.29
Individual Variance Share				
UVS	0.01	0.02	0.00	0.03
CVS	0.02	0.02	0.01	0.03
BVS	0.01	0.01	0.05	0.08
Industry Variance Share				
UVS	29.65	33.01	44.88	58.48
CVS	34.94	43.51	47.38	61.50
BVS	35.35	43.57	47.41	61.52

Notes: This table reports the results of a variance decomposition based on a estimates of the following regression:

$$c^g_{j,t} = X'_{jt}\beta^g + d'_t\delta_t + d'_s\delta_s + d'_i\delta_i + \varepsilon_{j,t} \tag{4.12}$$

The included individual characteristics, in X are age, gender, race and education dummies in addition to the generation dummies. See Section 4.4.1 for definitions of *UVS,CVS*, and *BVS*. All of the above numbers are percentages. We consider *standard controls* to be gender, race, age and education variables. *Generation* includes the generation dummies and *education* is comprised of the dummies relating to education outcome. All of the above numbers are percentages.

for the Silents to 62% for Millenials. This increasing role of industry is consistent with the literature on the sorting of workers across firms, for example Abowd et al. (1999) and Song et al. (2016). Song et al. (2016) show that two thirds of the rise of inequality since 1981 can be accounted for by the increased variance of earnings across firms.

To summarise, our variance decomposition analysis suggests little role for generational identity in determining real log wages, consistent with stagnating incomes. We also find that generation does explain around 3% of the variation in the labour share, but that consistent with the literature on inter-firm inequality this is much less than the growing role of industry as a determinant.

4.5 Hedonic Improvement

While working with real wages means we are allowing for the fact that the price level varies over time, it does not satisfactorily adjust for changes in quality. Thus, if the average loaf of bread (or phone) consumed by a member of the silent generation is qualitatively worse than that of the typical Millennial then our results will over-state the decline in living standards. This section asks what degree of hedonic progress we need assume for there to be equality across generations. We thus compute the equivalent variation in income necessary to make welfare equal across generations. Our approach is related to that of Jones and Klenow (2016) who compare welfare across countries by allowing for differences in hours worked, life-expectancy and inequality as well as consumption. They then obtain estimates of the equivalent consumption variation such that welfare is equalised across countries.

We begin by considering a very simple setting in which the median member of a given generation's utility in each period t depends only on their income y_t scaled by the relative quality level h_t. They, thus have a lifetime of income and, ignoring bequests, consumption such that $Y = \sum_t y_t$. Therefore we write lifetime utility as,

$$U = \sum_t u(y_t h_t) \tag{4.13}$$

Assuming log-utility, then we have that we can re-write lifetime consumption utility as,

$$U = \sum \ln y_t h_t = \sum \ln y_t + \sum \ln h_t \tag{4.14}$$

We observe $\sum_t \ln y_t^c$ for each cohort c. We can denote the total utility of the cohort with the highest lifecycle income $\sum \ln y_t^c$ as

$$U^* = \sum_t \ln y_t^* + \sum_t \ln h_t^* \tag{4.15}$$

If there is no intergenerational difference in consumption utility, then we simply have that $U^c = U^* \;\; \forall c$. Using this, and ignoring differences in life expectancy, we can then compute the implied relative quality level,

$$H^c = \frac{\bar{h}^c}{\bar{h}^*} = \frac{\sum_t h_t^c}{\sum_t h_t^*} = \exp[\ln \prod_t y_t^* - \ln \prod_t y_t^c] \tag{4.16}$$

Figure 4.8 plots the estimated value of H^c for both wages (left y-axis) and income (right y-axis) by year born. Given that, as we saw in Figure 4.1, the highest lifetime earnings are for those born around 1930 H_c is close to 1 for this cohort. Those born earlier were poorer and so they would have had to enjoyed a higher quality level to have the same level of welfare as the Silent Generation. Given the rapid rate of technological and societal progress between, say, 1900 and 1930 this seems implausible (Gordon, 2017).

To think through the limitations of this analysis, we consider the more general iso-elastic utility function used by King et al. (1988) and Mankiw and Weinzierl (2006),

$$U = \sum_t e^{-\rho t} \frac{(y_t h_o e^{gt})^{1-\gamma} e^{(1-\gamma)\nu(n_t)} - 1}{1-\gamma} \tag{4.17}$$

Figure 4.8: Hedonic Improvement for Total Population (log-utility)

Source: Current Population Survey (CPS)

Where $\nu(n_t)$ captures the disutility from working, and g is the rate of hedonic progress, and ρ is the subjective discount rate. Thus, generations' utilities will now depend on both the (total discounted) quality level of their consumption and the unpleasantness of their work as well as their total lifetime real incomes. This opens up two other potential sources of difference across generations. Firstly, discounting means that the shape of the life-cycle earnings trajectory (and of $\nu(n_t)$) matters. It also means that the (subjective) discount rate itself matters. Secondly, we now allow for the possibility that that those from later generations may not have consumed more over their lifetime, but may have endured less to get it.

The shape of the earnings trajectory was discussed in Section 4.2 in which we saw that generations subsequent to the Boomers both had lower average earnings and had to wait longer to attain maximum earnings. Thus, on this basis, and given a positive discount rate, the comparison in Figure 4.8 will understate the necessary rate of hedonic progress to equalise consumption for those born subsequent to the baby boom. On the other hand, increases in life-expectancy over the period may reduce discount rates which would offset this.

Thus, if, over their career, the median worker from the silent generation endured lower standards of labour protection or the benefits of fewer labour saving devices, or simply less meaningful work compared to their equivalent from Gen-X then the welfare comparison in Figure 4.8 is further complicated. As it will be, if on the other hand, similar jobs today are more stressful or tiring than in the past.

Such changes over time in the dis-utility of work are analysed by Kaplan and Schulhofer-Wohl (2018) who use data from the American Time-Use Study (ATUS) to compute average worker feelings about their jobs by occupation, and then analyse the implications of changes in the occupation structure of the US workforce to back out changes in aggregate feelings about work. As they discuss, in the absence of reliable and comparable historical data on workers' feelings about work, they need to assume that feelings by occupation have been stable overtime. This assumption, will be more plausible in some occupations than others, as some have changed little while others due to improvements in safety practices and technology are likely much better. In Figure 4.9 we plot for each of the six feelings they consider: Tiredness, Stress, Pain, Meaning, Happiness, and Sadness the average of these feelings (for available years) between the ages of 20 and 65 by year of birth. This then is the sum of feelings about each occupation weighted by the life-time share of that occupation out of total years worked for that birth cohort.

For all six feelings, the results suggest a consistent trend over time. But, these trends are not all in the same direction. Later generations experience more tiredness and more stress over their lifetime. But, one interpretation of the trends in other feelings is that work has become less important as a determinant of worker's feelings. It makes them less happy, but also less sad. It does not provide meaning, but also does not cause them pain. Such conflicting trends precludes an overall judgement about whether the disutility of work has increased or not. But, there is certainly no clear evidence that it has decreased. This in turn suggests that we can rule out changes in $\nu(n_t)$ across generations.

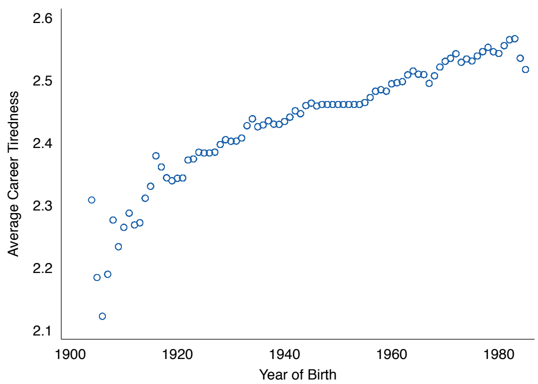
Overall, given this framework and the assumptions contained within we can draw the qualitative conclusions presented in Figure 4.8.

4.6 Conclusion

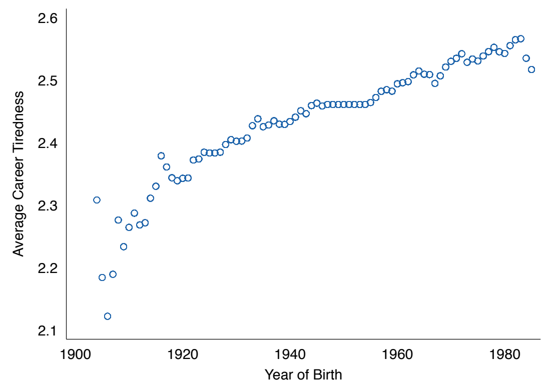
In this paper we document how, when comparing one generation to another, the median real wages of American men and women have been declining since the Silent generation born between 1925-1945. This is in contrast to consistent output and productivity growth over the same period (Jorgenson et al., 2008, Gordon, 2017). This phenomenon of *declining* incomes, first documented by Guvenen et al. (2017), is shown to be true conditional on a broad set of controls and allowing for unrestricted heterogeneity across industries. It has two key sets of implications. Firstly, this lack of intergenerational progress may, as argued by Friedman (2005), lead to an increasingly challenging environment for democracy. Secondly, given consistent productivity growth it implies that the labour share of income has been falling on a generational basis. We investigate this possibility and show that, within sector and state and conditional on a range of controls, that the labour share is

Figure 4.9: Dis-utility of Work by Year of Birth

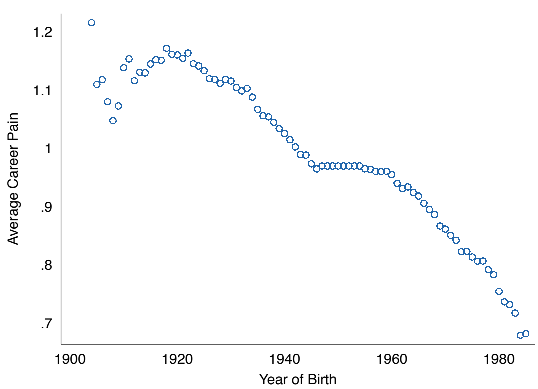
(a) Tired



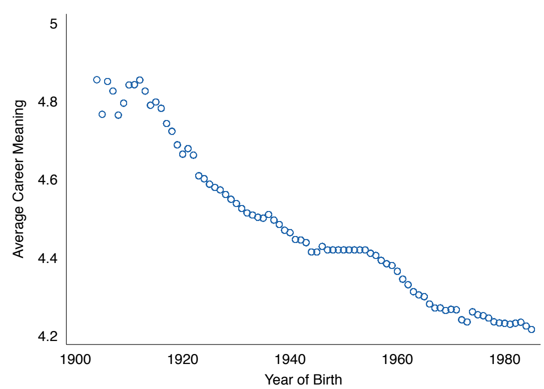
(b) Stressed



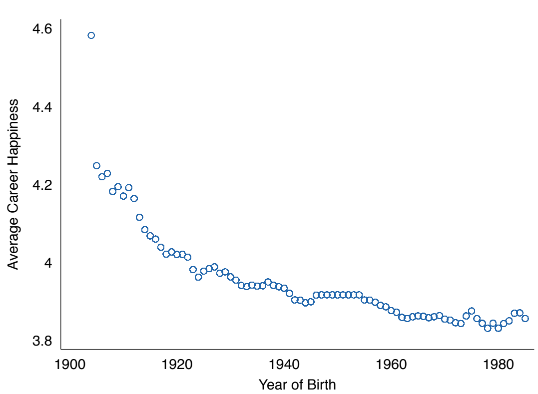
(c) Pain



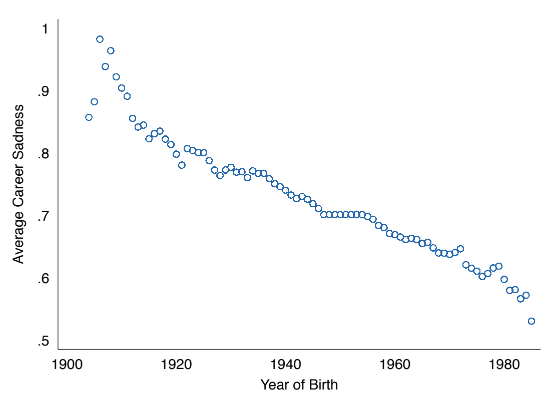
(d) Meaning



(e) Happiness



(f) Sadness



Source: Kaplan and Schulhofer-Wohl (2018) lower for Gen. X'ers and Millennials. Of course, to some degree this is implied by the findings of Karabarbounis and Neiman (2014), Piketty and Saez (2014), Autor and Salomons (2018), Autor et al. (2017b) that the labour share is decreasing, but what is novel is that the labour share is systematically different to workers of different generations, even conditioning on age and year, etc. What it is that has caused this change is not something this paper speaks to, but we note that the change is consistent across most industries and for all sub-groups. Furthermore, it is not immediately obvious which of the leading explanations such as the rise of 'superstar firms' (Autor et al., 2017b), automation Autor and Salomons (2018), the price of investment goods (Karabarbounis and Neiman, 2014), or the rate of productivity growth (Grossman et al., 2018), or the rise off-shorting (Elsby et al., 2013), would predict such large changes between generations, other things equal. Our variance decomposition analysis in Section 4.4 showed that these changes are of quantitative importance even once we allow for the role of state, industry, education, and demographics. The final part of the paper considers whether hedonic improvements in the quality of goods and services have been sufficiently large to imply improving standards of living across generation. We find that the implied qualitative improvement is around 30% which is a substantial improvement, but not necessarily an implausible one.

Appendix A

Chapter 2

A.1 Data Appendix

The Panel Study of Income Dynamics (PSID) is a longitudinal survey which follows families over time. The survey ran annually from 1968 to 1997 and biennially from 1997 until 2015. The survey was initially sampled almost 3000 families which were representative of the US demographics (The SRC Sample), and an oversampling of low income households which comprised of around 2000 families (The Survey of Economic Opportunity, or SEO Sample). The sample members here were continued to be surveyed, even if they moved out and began their own family unit.

Following this, there was a number of refresher samples and updates to the core set of families. This included 2000 latino families in 1990 and 500 immigrant families in 1997/99. Due to the following of individuals in families and the creation of new family units, the number of households and individuals interviewed for the PSID has grown substantially since the first wave in 1968. In the 2015 wave there was almost 10,000 families which were interviewed.

The nature of the PSID and its core property of following families means that there is substantial micro data available about different generations within a family. Thus, the PSID is an ideal data set for studying intergenerational transmissions as one is able to observe both parent and child (plus other family members, such as siblings) at different points in time.

Because of our interest in the intergenerational links, we limit our analysis to the period since 1980. However, we make use of the entire dataset to construct the relevant variables for parents. The SEO sampled are excluded from the core of our results.[1] As including them would mean the oversampling of the bottom of the income distribution.

To classify the children into their socio economic group, we take the average of the family income when they were aged between 14 and 18. We then classify the bottom quartile and top quartile as low and high socio economic status respectively. Hence those in the second and third quartile are considered middle socio economic status.

The outcome of interested will be the cumulative distribution function of children's labour income, that is solely their earnings from employment. To limit the number of groups for which we need estimate distributions for, we create a number of discrete variables for our controls. Firstly, we consider two education groups; those with high school or less and those with at least some college. We expand this further to include, has a college degree or not, and a dummy for graduating high school. Secondly, we group age into young ($18 \leq age \leq 35$), middle ($35 < age \leq 50$) and older ($50 \leq age \leq 65$), dropping anyone who cannot be classified into these groups. Our findings are robust to different classifications of our controls. An additional measure we use is to pool years in the pre 1997 data.

[1] This has no bearing on the qualitative results. Hence it is not important whether they are included for our core findings.

Table A.1: Summary Statistics including SEO Sample

	Overall		*Low Income*		*Middle Income*		*High Income*	
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
Children								
Female	0.55	0.50	0.59	0.49	0.54	0.50	0.53	0.50
Age	35.08	7.48	34.97	7.35	35.11	7.48	35.12	7.60
White	0.63	0.48	0.28	0.45	0.66	0.47	0.90	0.30
Black	0.35	0.48	0.70	0.46	0.31	0.46	0.08	0.27
Hispanic	0.01	0.09	0.01	0.08	0.01	0.10	0.00	0.05
Other Ethnicity	0.01	0.12	0.02	0.12	0.01	0.12	0.01	0.12
Years of Education	13.25	2.09	12.36	1.92	13.11	1.96	14.44	1.94
High School Graduate	0.90	0.30	0.81	0.39	0.91	0.29	0.97	0.18
College Graduate	0.24	0.43	0.10	0.30	0.20	0.40	0.46	0.50
In Employment	0.80	0.40	0.74	0.44	0.82	0.39	0.84	0.36
Labour Income	28,184	41,325	18,927	18,462	26,860	40,215	39,902	54,826
Mother								
White	0.36	0.48	0.15	0.35	0.38	0.49	0.55	0.50
Black	0.24	0.43	0.47	0.50	0.21	0.41	0.05	0.23
Hispanic	0.01	0.08	0.01	0.08	0.01	0.09	0.00	0.03
Other Ethnicity	0.39	0.49	0.38	0.48	0.40	0.49	0.40	0.49
High School Graduate	0.60	0.49	0.30	0.46	0.60	0.49	0.90	0.31
College Graduate	0.09	0.28	0.01	0.11	0.05	0.22	0.23	0.42
Average Hours Mother Worked per Year	955	812	743	765	1,008	805	1,069	833
Father								
White	0.33	0.47	0.09	0.29	0.34	0.47	0.54	0.50
Black	0.14	0.35	0.19	0.40	0.16	0.37	0.05	0.22
Hispanic	0.01	0.09	0.01	0.09	0.01	0.09	0.01	0.08
Other Ethnicity	0.53	0.50	0.71	0.46	0.49	0.50	0.41	0.49
High School Graduate	0.58	0.49	0.22	0.42	0.52	0.50	0.87	0.34
College Graduate	0.17	0.38	0.01	0.10	0.08	0.27	0.41	0.49
Average Hours Father Worked per Year	2,054	776	1,403	1,004	2,059	707	2,369	515
Family								
Average Family Income	50,245	39,746	16,898	5,396	43,431	11,415	97,215	51,154
Observations	135,354		33,850		67,656		33,848	

Table A.1 contains the summary statistics which include the Survey of Economic Opportunity, a sample in the PSID of low income households. We exclude this from the main findings in the paper however this table supports the claims that our qualitative results would be robust to its inclusion. The inclusion of the sample shifts the classification of the parents, so if anything, the inclusion of the sample would suggest stronger findings than that which are presented.

A.2 Variance of the Bounds

We can derive the variance of the bounds similarly to how bounds to the distribution are derived under various assumptions. Recall that we observe wages in the case where an individual is working, $W = 1$, with a probability of $P(x)$. We know that the variance of the cumulative distribution function $F(y|x)$, takes the following form,

$$Var(\widehat{F}(y|x)) = \frac{1}{N}\widehat{F}(y|x)[1 - \widehat{F}(y|x)]$$

Following Imbens and Manski (2004), we denote σ^2 as the variance of the cumulative distribution function in the case where we observe the wages. Formally, $\sigma^2 = Var(\widehat{F}(y|x, W = 1))$. Using this and the assumptions on the bounds we can estimate the variance for the lower and upper bound (denoted $\hat{\sigma}_l^2, \hat{\sigma}_u^2$ respectively) given our set of assumptions.

Beginning with our worst case bounds; recall these are the bounds with minimal assumptions. We derive, using the law of total variance, these to be,

$$\hat{\sigma}_l^2 = \hat{\sigma}^2 \cdot P(x) + P(x) \cdot [1 - P(x)] \cdot [\widehat{F}(y|x, W = 1)]^2$$

$$\begin{aligned}\hat{\sigma}_u^2 = \hat{\sigma}^2 \cdot P(x) + P(x) \cdot [1 - P(x)] \cdot [\widehat{F}(y|x, W = 1)]^2 \\ - 2\widehat{F}(y|x, W = 1) \cdot P(x) \cdot [1 - P(x)] + P(x) \cdot [1 - P(x)]\end{aligned}$$

Additional assumptions we make, such as stochastic dominance and the median restriction imply restrictions on the lower bound only. For this reason the estimate for the variance of the upper bound will remain unchanged for any set of assumptions and thus only the lower bound will need to be updated. For the stochastic dominance assumption the lower bound variance will simply be the estimate for the variance of the CDF. That is,

$$\hat{\sigma}_l^2 = \frac{1}{N}\widehat{F}(y|x, W = 1)[1 - \widehat{F}(y|x, W = 1)]$$

Where we introduce the median restriction, the lower bound variance will become,

$$\hat{\sigma}_l^2 = \hat{\sigma}^2 \cdot P(x) + P(x) \cdot [1 - P(x)] \cdot [\widehat{F}(y|x, W = 1)]^2 - \\ \widehat{F}(y|x, W = 1) \cdot P(x) \cdot [1 - P(x)] + \frac{1}{4}(1 - (P(x))^2)$$

A.3 Asymptotic Variance

The asymptotic variance of the interquartile range,

$$\sqrt{n}(IQR - (\eta_{0.75} - \eta_{0.25})) \longrightarrow N\left(0, \frac{1}{16}\left[\frac{3}{f^2(\eta_{0.75})} + \frac{3}{f^2(\eta_{0.25})} - \frac{2}{f(\eta_{0.75})(\eta_{0.25})}\right]\right)$$

where $\eta_p = F^{-1}(P)$

The asymptotic variance of the median,

$$Var(Median) = \frac{1}{4\left\{f_x(\theta)^2\right\}}$$

where $f_x(\theta)$ if the probability density function which corresponds the the cumulative distribution function.

A.4 Results Appendix

Table A.2: Bounds to the Change in the Interquartile Range.

	Low Income	Middle Income	High Income
Worst Case	-2.55, 1.30	-0.86, 0.89	-0.52, 0.53
	[-2.62, 1.36]	[-0.91, 0.93]	[-0.55, 0.57]
Median Restriction	-2.25, 1.00	-0.67, 0.60	-0.42, 0.33
	[-2.33, 1.06]	[-0.72, 0.63]	[-0.46, 0.37]
Stochastic Dominance	-1.89, 0.50	-0.48, 0.45	-0.32, 0.21
	[-1.96, 0.57]	[-0.53, 0.48]	[-0.35, 0.24]

Note: The 90% Confidence Interval for the change is given in square brackets

Table A.3: Bounds to the Change in Interquartile Range by Gender

	Low Income	Middle Income	High Income
Son's			
Worst Case	-0.18, 0.57	0.07, 0.52	0.12, 0.42
	[-0.21, 0.60]	[0.05, 0.54]	[0.10, 0.44]
Median Restriction	-0.14, 0.47	0.09, 0.43	0.14, 0.38
	[-0.17, 0.50]	[0.07, 0.44]	[0.12, 0.40]
Stochastic Dominance	-0.03, 0.38	0.14, 0.37	0.17, 0.33
	[-0.06, 0.40]	[0.12, 0.39]	[0.15, 0.35]
Daughter's			
Worst Case	-3.68, 3.30	-2.47, 1.07	-3.33, 0.95
	[-3.83, 3.59]	[-2.51, 1.67]	[-3.38, 1.79]
Median Restriction	-3.04, 2.73	-2.15, 0.87	-2.92, 0.76
	[-3.18, 3.03]	[-2.20, 1.48]	[-2.97, 1.59]
Stochastic Dominance	-2.67, 2.22	-1.89, 0.42	-2.66, 0.51
	[-2.80, 2.52]	[-1.94, 1.12]	[-2.70, 1.25]

Note: The 90% Confidence Interval for the change is given in square brackets

Figure A.1: Bounds to the Median Wage by Son's and Daughter's with no College.

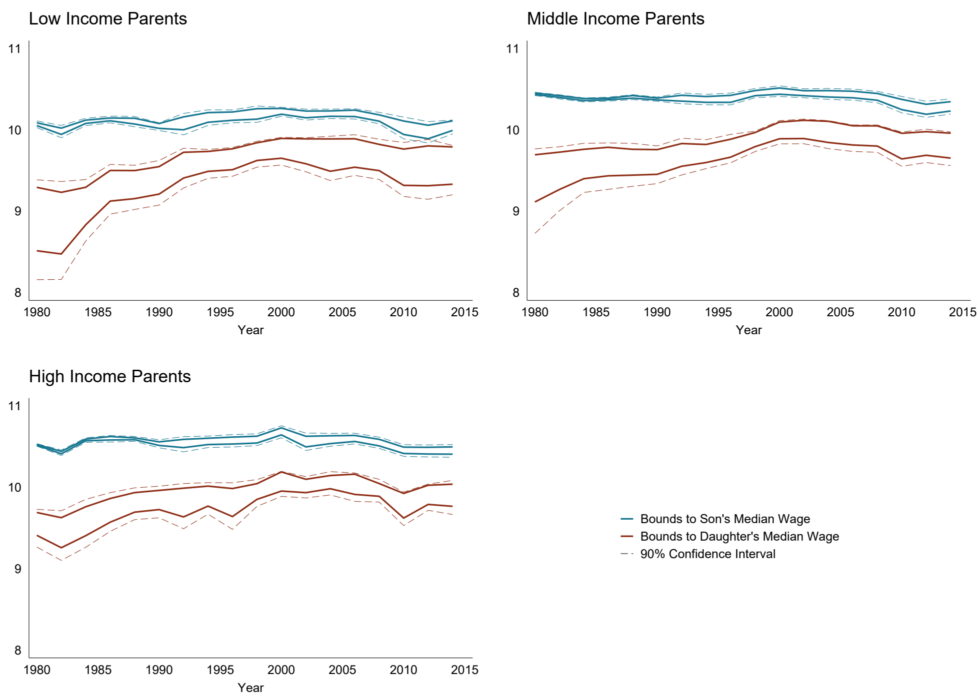


Figure Notes: On the vertical axis is the log median annual wage by year on the horizontal axis. Estimates are produced with the median restriction assumption along with the 90% Confidence intervals.

Table A.4: Bounds to the Change in Median Wage by Gender and Education

	Low Income	Middle Income	High Income
Son's with No College			
Worst Case	-0.13, 0.12	-0.23, -0.05	-0.12, 0.01
	[-0.14, 0.14]	[-0.24, -0.04]	[-0.14, 0.02]
Median Restriction	-0.09, 0.06	-0.22, -0.09	-0.12, -0.02
	[-0.11, 0.07]	[-0.23, -0.08]	[-0.13, -0.01]
Monotonicity (IV)	-0.05, 0.06	-0.20, -0.08	-0.02, 0.02
	[-0.58, 0.91]	[-0.52, 0.44]	[-0.61, 0.93]
Stochastic Dominance	-0.09, 0.03	-0.22, -0.13	-0.12, -0.04
	[-0.11, 0.05]	[-0.23, -0.11]	[-0.13, -0.03]
Daughter's with No College			
Worst Case	-0.35, 1.51	-0.36, 0.96	-0.09, 0.80
	[-0.38, 1.56]	[-0.39, 1.00]	[-0.11, 0.83]
Median Restriction	0.04, 1.28	-0.04, 0.85	0.08, 0.63
	[0.02, 1.33]	[-0.06, 0.88]	[0.06, 0.65]
Monotonicity (IV)	0.00, 0.65	-0.17, 0.61	0.21, 0.54
	p-0.58, 0.91]	[-0.52, 0.44]	[-0.61, 0.93]
Stochastic Dominance	0.05, 1.18	-0.02, 0.76	0.10, 0.61
	[0.03, 1.23]	[-0.04, 0.80]	[0.08, 0.64]
Son's with College			
Worst Case	-0.07, -0.02	0.07, 0.18	0.11, 0.19
	[-0.08, -0.01]	[0.05, 0.20]	[0.09, 0.20]
Median Restriction	-0.06. -0.02	0.07, 0.17	0.11, 0.17
	[-0.07, -0.01]	[0.06, 0.18]	[0.10, 0.18]
Monotonicity (IV)	-0.01, -0.01	0.08, 0.17	0.14, 0.19
	[-0.33, 1.34]	[-0.16, 0.70]	[0.07, 0.91]
Stochastic Dominance	-0.06, -0.03	0.08, 0.14	0.12, 0.16
	[-0.07, -0.02]	[0.06, 0.15]	[0.10, 0.17]
Daughter's with College			
Worst Case	0.17, 0.75	0.12, 0.58	0.24, 0.61
	[0.15, 0.77]	[0.11, 0.60]	[0.23, 0.62]
Median Restriction	0.37, 0.69	0.27, 0.55	0.31, 0.53
	[0.36, 0.71]	[0.26, 0.57]	[0.30, 0.54]
Monotonicity (IV)	0.35, 0.40	0.32, 0.42	0.37, 0.45
	[-0.33, 1.34]	[-0.16, 0.70]	[0.07, 0.91]
Stochastic Dominance	0.37, 0.66	0.27, 0.53	0.32, 0.51
	[0.35, 0.68]	[0.26, 0.55]	[0.31, 0.52]

Note: The 90% Confidence Interval for the change is given in square brackets

Figure A.2: Bounds to the Change in Interquartile Range by Gender and Education.

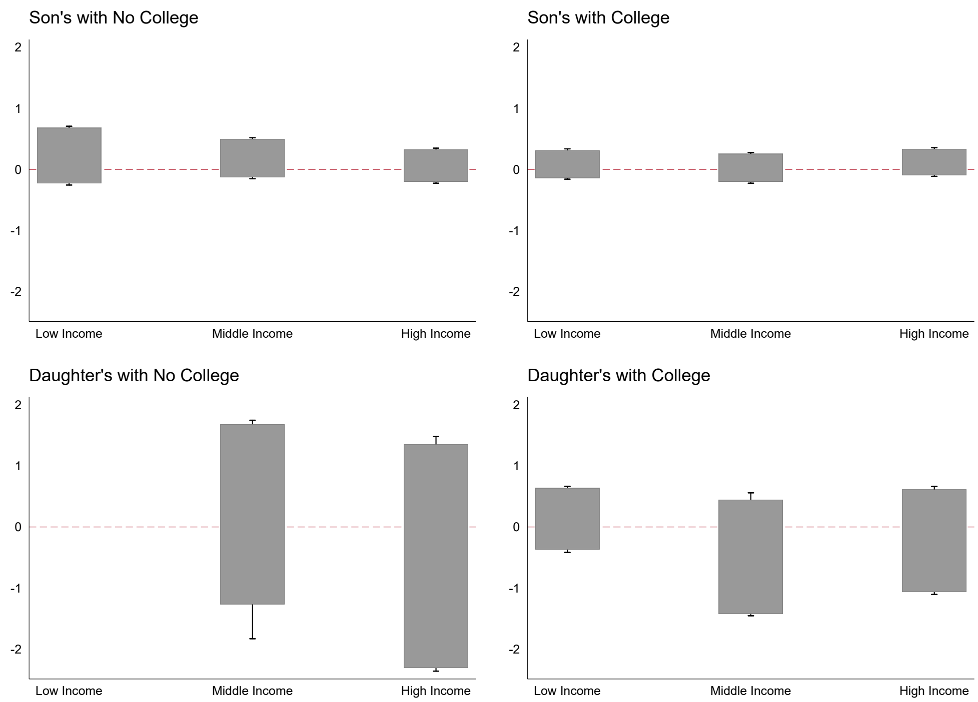


Figure Notes: On the vertical axis is the change in interquartile range between 1988 and 2014. Estimates are produced with the median restriction assumption, the limits of the box refer to the lower and upper bound to the change and the whiskers to the 90% confidence interval. The result for low income daughters with no college is not defined and is so omitted from the graph.

Appendix B

Chapter 3

B.1 Proof of Proposition 1

Proof of Proposition 1. Focusing on income inequality and following Milanovic (1997) we can write the Gini Coefficient of Income as:

$$\theta(W) = \frac{1}{\sqrt{3}}\frac{\sigma_W}{\overline{W}}\rho(W, r_W)\frac{\sqrt{N^2-1}}{N} \approxeq \frac{1}{\sqrt{3}}\frac{\sigma_W}{\overline{W}}\rho(W, r_W),$$

where $\overline{W}$, σ_W are the mean and standard deviation of individual income W, r_W is the rank of a specific income level W and $\rho(W, r_W)$ is the correlation of W with its rank r_W. To proceed, observe that $\rho(W, r_W) \in [0, 1]$ and that $\rho(W, r_W) = 0$ if and only if $W = \overline{W}\,\forall\, W$, otherwise $\rho(W, r_W) \in (0, 1]$. In combination with the fact that $\sigma_W \geq 0$ but also $\sigma_W = 0$ if and only if $W = \overline{W}\,\forall\, W$, implies that as longs as the set $W \neq \overline{W}$ is non-empty $\theta(W) > 0$. Results for the Gini Coefficient of Wealth can be established with the same arguments. $\square$

B.1.1 Data Appendix

Current Population Survey

The Current Population Survey (CPS) has been conducted monthly by the U.S. Census Bureau, since 1962. In what follows we outline the nature of the survey and our treatment of the data. This treatment has been closely informed by those of Heathcote et al. (2010), and where possible we have done exactly as they did. Indeed, one important contribution of their paper was to establish a treatment of the data that provided estimates that could be cross-validated against those from the Panel Study of Income Dynamics (PSID) and the Consumer Expenditure Survey (CEX).

The CPS surveys a representative sample of each state population restricted to those over the age of 15 and who are not in the armed forces nor any kind of institution such as a prison or hospice. In total it surveys around 60,000 households each month. Households are sampled using a $4-8-4$ sampling scheme, in which households are interviewed for four consecutive months, not visited for eight months, and then surveyed again for four more consecutive months at the same time the following year. Most important for our purposes is the data collected in the March Annual Social and Economic Supplement (ASEC). This cross sectional annual supplement contains detailed data relating to income and employment.

All of our estimates are produced using the March ASEC weights which correspond to individual level observations. We first restrict our sample by dropping the small number of observations for which 'bad', i.e. negative weights are recorded, although this does not affect our results. Secondly, we remove individuals younger than age 18 and older than age 78 when using total income measures. When we consider labour income inequality the age range included is 18 to 65.

The CPS data are top-coded and this might lead us to understate inequality. In our preferred results we do not use any correction for top-coding but we obtain the same results if we instead apply the Pareto-interpolation correction suggested by Heathcote et al. (2010)[1] More important for our analysis is the slight discrepancy between the survey year and the year to which the survey refers. Given the retrospective nature of the survey we assign values from the survey in year t to calendar year $t-1$. That is, for example, results for 2002, are based on the 2003 survey which was conducted in March that year.

The two income variables we are interested in are, again like Heathcote et al. (2010), labour income and total income. Our labour income variable is each respondent's total pre-tax wage income from employment. The total income variable records the total, pre-tax, personal income or losses from all sources. Both variables are adjusted for inflation using the CPI-U series of the Bureau of Labor Statistics.

Perhaps the most substantive decision is how to handle missing data. Data can be missing either because a household did not respond, or because a particular question was not answered. Weights are used to address the former problem, and "hot-deck" imputation (assigning the response from a randomly chosen statistically similar household). We, again, follow Heathcote et al. (2010) and retain these imputed values and use the CPS provided survey weights.

Luxembourg Income Study Database (LIS)

The Luxembourg Income Study (LIS) provides a harmonised data set of microdata recording a broad range of economic and demographic characteristics drawn from various nationally representative surveys. Data are compiled at both the individual

[1]This correction assumes that underlying distribution of income has a Pareto distribution. By estimating the parameter of this Pareto distribution from the non-top-coded upper end of the distribution, allows estimation of the true mean of the top-coded incomes.

and household levels. For each wave, from each country, LIS takes data for the individual and the household level, with variables relating to socio-demographics, household characteristics, labour market and flow variables. The individual file is made up of the members of the households included in the household level files, where their individual observations regarding income and expenditure are summed to create the household aggregate information. For our purposes we use the individual level income data only.

The harmonisation procedure involves two main components. Firstly, ensuring the variables are comparable in terms of their definitions and in the coding convention applied, for example with respect to categorical variables. Secondly, missing values are processed to ensure both a consistent coding across countries and waves, but also given the differing questions asked by each national survey-wave where possible missing data are derived from the available data. For example, if the underlying survey does not contain information about unemployment but does contain sufficient employment data then unemployment data is derived appropriately.

The datasets produced by LIS are representative of the total population of that country for the given year. To this end the most appropriate weights provided by the original surveys are selected, and where necessary missing individual or household level weights are derived using the provided weighting data. The key criteria for the choice of weight variable, is that they deliver nationally representative results and in the cases where there is a choice of these priority is given to those which are designed to accurately capture the population income distribution.

We consider two main income variables from the LIS datasets taken from the individual level data files. These values are corrected for inflation by LIS using the Consumer Price Index (CPI).

Personal Monetary Income This is the total monetary income that an individual receives from labour and transfers. As such it is akin to the pre-tax total income in the CPS, and we will refer to it as Total Income.

Labour Monetary Income Labour income includes any monetary payments received from employment, in addition any profits or losses accruing from self employment.

We can additionally consider both the value monetary and non-monetary income however not all data sets are as good as reporting non-monetary income so this component maybe under reported in many cases. Regardless of this difference we can find similar results for both monetary and non-monetary incomes. We limit the age range consider to 18-78 when using personal monetary income, and to 18-65 for labour monetary income.

The LIS classifies each data set depending on the kind of income that the host data provider report. These groups are either *gross*, *net*, or *mixed*. A majority of the datasets are *gross*, that is the income amounts reported are gross of income taxes and social security employer contributions. This is contrasted to the *net* datasets which there is no information provided regarding taxes and other contributions. Finally, *mixed* datasets where that taxes and contribution data is not sufficiently available to be purely classified as either *gross* or *net*.

Luxembourg Wealth Study (LWS)

Our estimates of wealth inequality use data from the Luxembourg Wealth Study Database (LWS) . This combines representative national surveys on the basis of the same principles as the LIS, producing harmonised cross country data. A key difference is that wealth variables are measured at the level of the household unit.

Therefore, we need to assign an 'age' to each household to calculate *natural* and *adjusted* inequality. To do so, we use the age of the head of household. This choice is unimportant for our results. All of our estimates are produced using the weights provided by LWS, and we allow net wealth to be negative. Wealth data are often top-coded and the wealthy are often oversampled due to higher rates of non-response. This can mean, given the small number of very wealth individuals, that results may not be truly representative. To address bias due to this we drop the top 1% of wealth observations in each country. Data for the United States are drawn from the Survey of Consumer Finances (SCF) and so we follow the approach of Heathcote et al. (2010) who trim the SCF so that the mean income is consistent across all their datasets.

We choose disposable net worth (non-financial assets plus financial assets (excluding pensions) minus total liabilities) as our measure of wealth. A driving factor in this choice is the inconstistent way in which pension wealth is measured across countries and in some cases not available in the LWS dataset. So for this reason we have decided not to use the measure of wealth which includes pensions.

B.1.2 Future Cohorts Simulation Procedure

In order to simulate the future cohort shares we create a Leslie Matrix (Leslie, 1945, 1948), a form of projection matrix that is a standard tool in Mathematical Demography. We have information regarding the population cohort sizes for time t, but we are interested in forecasting the population for time $t+s$. Given we have data, for each age i on age specific fertility rates β_i and death rates μ_i we can construct the Leslie matrix, $\mathbf{L}$ which has age specific fertility on its top row, and age specific survival rates on the first subdiagonal. Multiplying the vector of cohort population shares (ordered by age) for year t P_t by $\mathbf{L}$ gives the vector of population shares for the subsequent year P_{t+1}. That is $P_{t+1} = \mathbf{L}P_t$, where:

$$\mathbf{L} = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \dots & \beta_w \\ 1-\mu_0 & 0 & 0 & \dots & 0 \\ 0 & 1-\mu_1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix}.$$

Where subscript w denotes the maximum possible attainable age. P_t is the vector of the current population cohort sizes ordered by age. Thus the population in year $t + s$ is obtained by calculating:

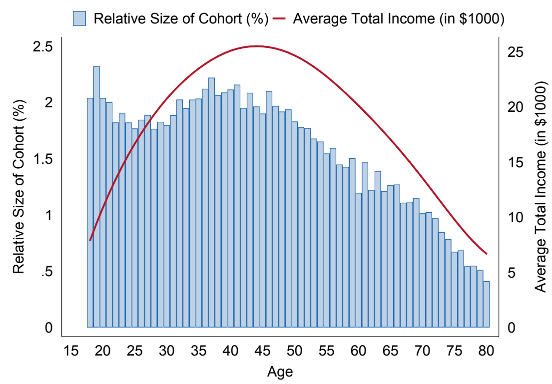
$$P_{t+s} = \mathbf{L}^s P_t. \tag{B.1}$$

Performing this procedure for each horizon $s \in 1, \dots, 40$ us our population forecasts and maps the transition of the population returning to its long run steady-state following the shock constituted by the Baby Boom.[2] Figure 18 excludes Austria, Spain, Italy and Hungary as the data sets used for the simulations are all *gross*, unlike the available historical data for these countries.

[2]We are grateful to Timo Trimborn for sharing his code for this procedure.

B.2 Additional Results

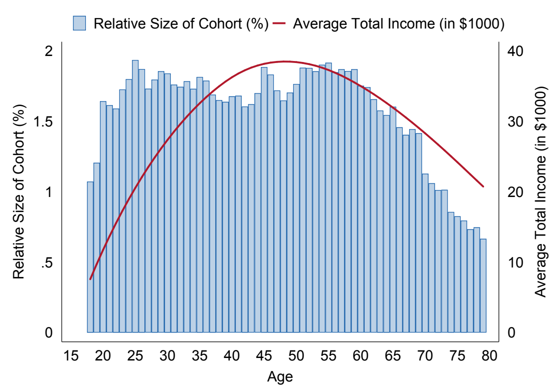
Figure B.1: Income and cohort size by age group United States : 1961



Source: Authors' calculations using LIS data.

Notes: We consider men who are aged 18-78 for total income and who have positive earnings. Results are calculated using individual level weights.

Figure B.2: Income and cohort size by age group United States : 2015



Source: Authors' calculations using LIS data.

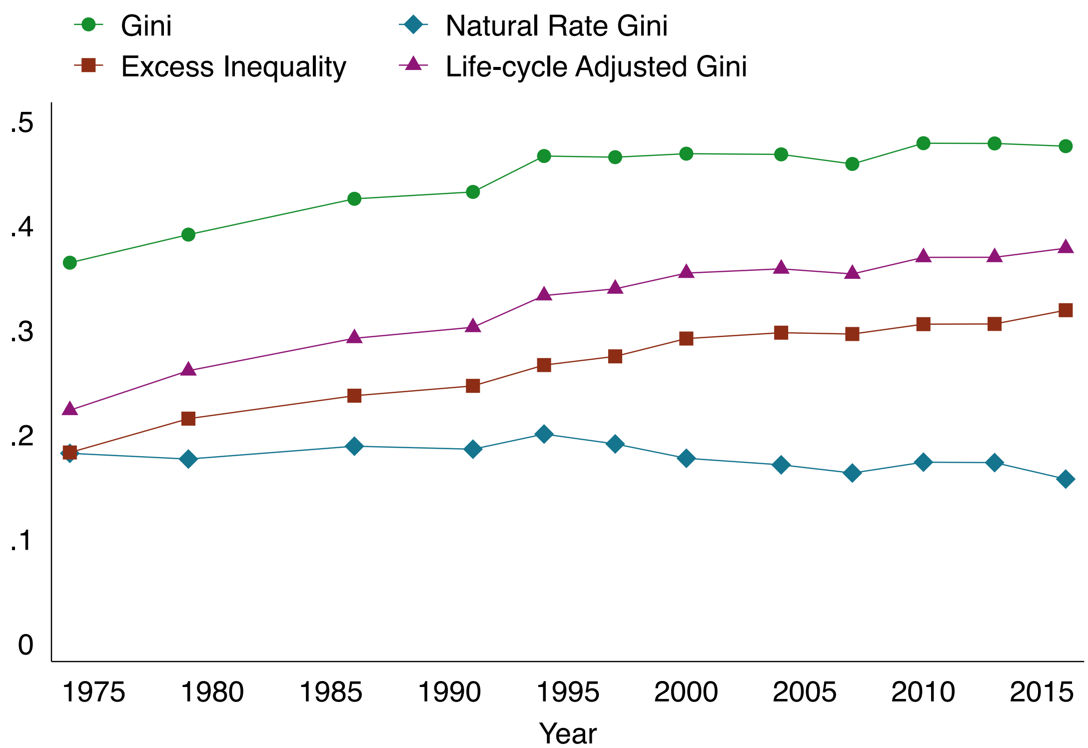
Notes: We consider men who are aged 18-78 for total income and who have positive earnings. Results are calculated using individual level weights.

Figure B.3: Adjusted and Unadjusted Gini of Total Income for the United States using LIS: 1974 - 2016

Source: Authors' calculations using LIS data.

Notes: We consider men who are aged 18-78 for total income and who have positive earnings. Results are calculated using individual level weights.

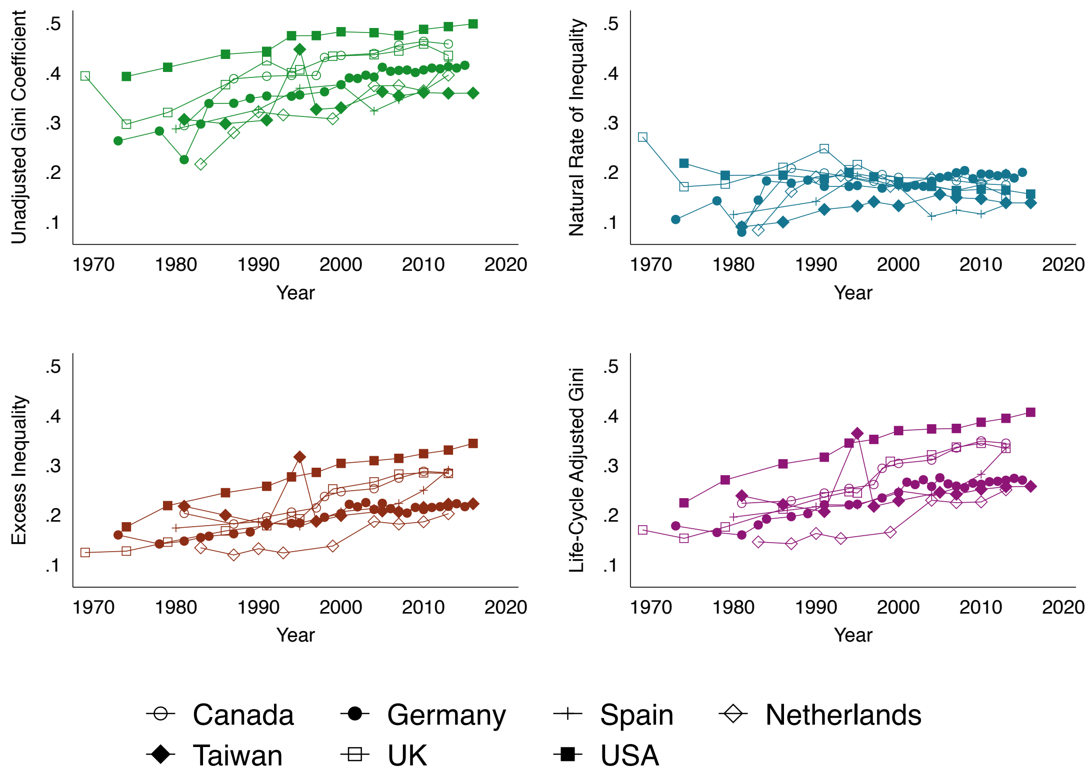
Figure B.4: Adjusted and Unadjusted Gini of Labour Income for the United States using LIS: 1974 - 2016



Source: Authors' calculations using LIS data.

Notes: We consider men who are aged 18-65 for labour income and who have positive earnings. Results are calculated using individual level weights.

Figure B.5: Adjusted and Unadjusted Gini of Total Income: Selected Countries: 1969-2016



Source: Authors' calculations using LIS data.

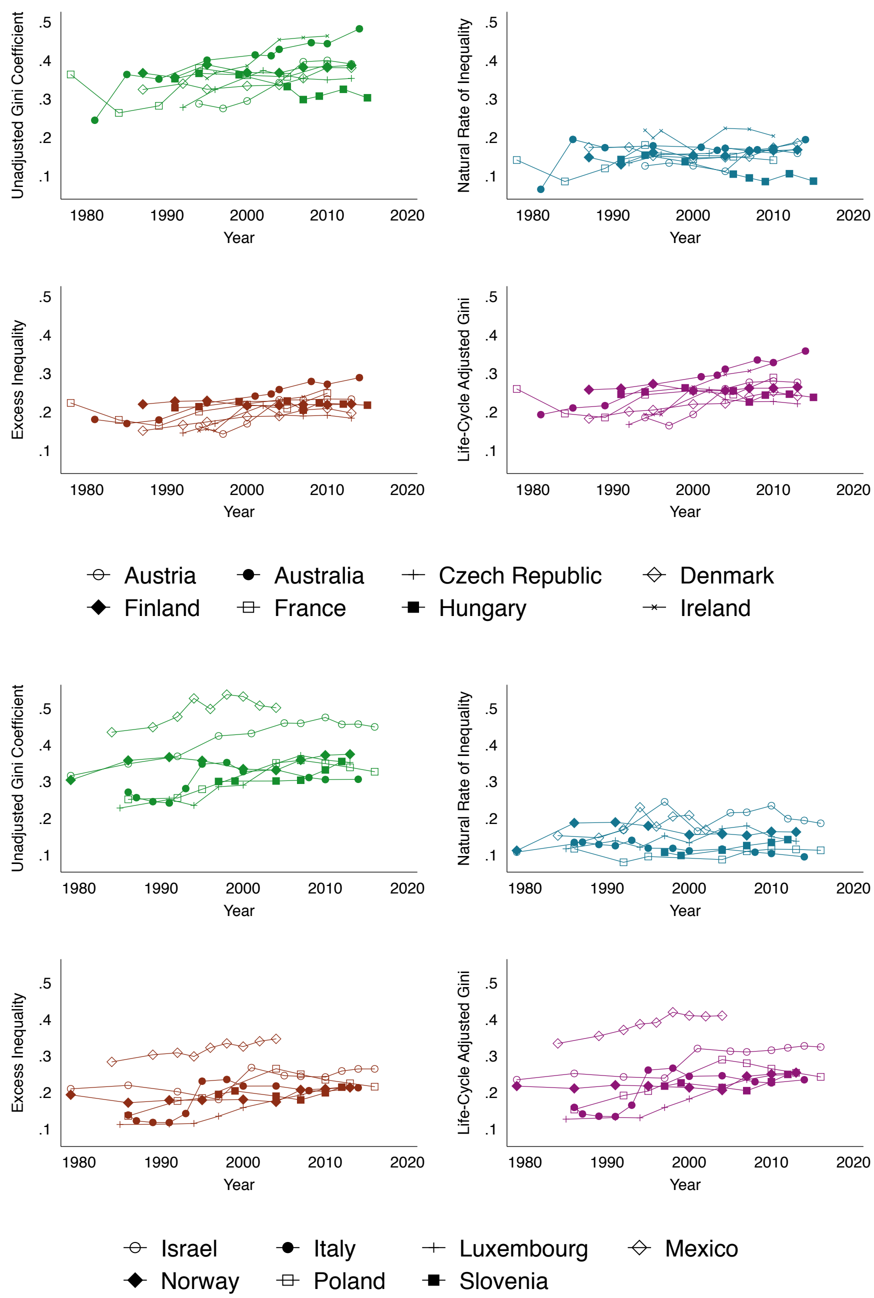
Notes: All results are calculated using data on gross incomes with the exception of Spain which are net incomes (with exception of wave IX). We consider ages 18-78 for total income and who have positive earnings. Results are calculated using individual level weights.

Table B.1: Country Specific Trend Estimates

	Actual		*Adjusted*		
Country	Total	Labour	Total	Labour	N
Austria	0.74***	0.80***	0.67***	0.75***	7
	(0.13)	(0.13)	(0.12)	(0.14)	
Australia	0.55***	0.41***	0.51***	0.35***	10
	(0.12)	(0.05)	(0.02)	(0.02)	
Canada	0.27***	0.52***	0.24***	0.48***	11
	(0.06)	(0.10)	(0.06)	(0.03)	
Czech Republic	0.31*	0.41 ***	0.23	0.22**	7
	(0.14)	(0.07)	(0.12)	(0.09)	
Germany	0.39**	0.44***	0.29***	0.32***	27
	(0.04)	(0.04)	(0.02)	(0.02)	
Denmark	0.06	0.23***	0.07**	0.20***	8
	(0.05)	(0.05)	(0.02)	(0.03)	
Spain	0.32**	0.31**	0.39***	0.34**	8
	(0.09)	(0.13)	(0.07)	(0.12)	
Finland	-0.01	0.05	-0.05***	-0.06	8
	(0.02)	(0.05)	(0.01)	(0.05)	
France	0.17	0.33**	0.10	0.24	7
	(0.19)	(0.13)	(0.14)	(0.15)	
Hungary	-0.27***	-0.39***	-0.06	-0.27***	8
	(0.06)	(0.07)	(0.03)	(0.06)	
Ireland	0.75***	0.70***	0.92***	0.83***	7
	(0.09)	(0.09)	(0.07)	(0.07)	
Israel	0.41***	0.43***	0.29***	0.26***	11
	(0.05)	(0.05)	(0.03)	(0.03)	
Italy	0.29***	0.29***	0.52***	0.27***	12
	(0.08)	(0.08)	(0.09)	(0.07)	
Luxembourg	0.51***	0.61***	0.53***	0.57***	9
	(0.09)	(0.07)	(0.04)	(0.06)	
Mexico	0.59***	0.40**	0.62***	0.40***	9
	(0.12)	(0.13)	(0.07)	(0.07)	
Netherlands	0.36***	0.62***	0.35***	0.45***	9
	(0.09)	(0.05)	(0.05)	(0.05)	
Norway	-0.15**	0.27***	-0.21**	0.19***	9
	(0.05)	(0.06)	(0.07)	(0.02)	
Poland	0.35***	0.36***	0.35**	0.30**	8
	(0.08)	(0.09)	(0.10)	(0.08)	
Slovenia	0.32**	0.30*	0.08	0.16	6
	(0.07)	(0.14)	0.10	0.10	
Taiwan	0.16**	0.14***	0.05	0.13*	11
	(0.07)	(0.04)	(0.07)	(0.07)	
United Kingdom	0.28**	0.50***	0.48***	0.51***	12
	(0.09)	(0.03)	(0.04)	(0.03)	
United States	0.24***	0.25***	0.40***	0.35***	12
	(0.02)	(0.04)	(0.04)	(0.04)	

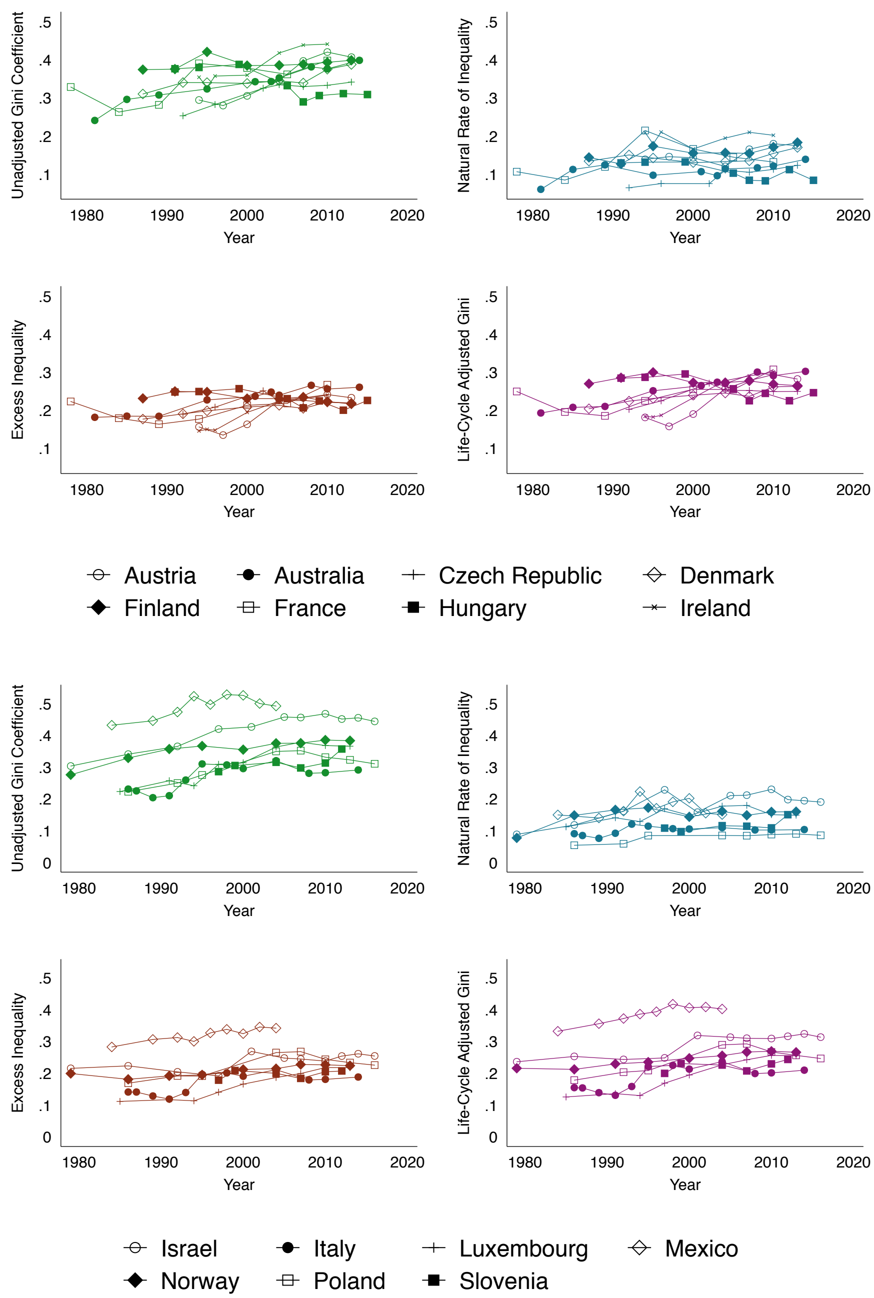
Coefficients are country specific time trends obtained using the Mean Group estimator of Pesaran and Smith (1995). See Table 3.1 for further details.

Figure B.6: LIS Additional Countries, Total Income



Source: Authors' calculations using LIS data.
Notes: These are the countries for which a sufficient time series is available not reported in Figure 3.8. Note that, however, data for these other countries are not consistently classified as gross or net. Most datasets are classified as Gross. France is all classed as mixed and Slovenia is classed as Net. Austria, Belgium, Hungary, Israel, Italy, Luxembourg and Poland do not have a consistent classification over the time series. All others are for gross income. We consider Men aged between 18-78 and who have positive income. Results are calculated using individual level weights.

Figure B.7: LIS Additional Countries, Labour Income



Source: Authors' calculations using LIS data.
Notes: These are the countries for which a sufficient time series is available not reported in Figure 3.8. Note that, however, data for these other countries are not consistently classified as gross or net. Most datasets are classified as Gross. France is all classed as mixed and Slovenia is classed as Net. Austria, Belgium, Hungary, Israel, Italy, Luxembourg and Poland do not have a consistent classification over the time series. All others are for gross income. We consider Men aged between 18-65 and who have positive income. Results are calculated using individual level weights.

Appendix C

Chapter 4

C.1 Data Appendix

C.1.1 Current Population Survey (CPS)

The CPS is individual micro level data which is available from 1962 to 2017. With the sample weights it is representative of the US population each year. The core of our analysis is at the cohort or generation level. We discuss how we define these groups below before going on to discuss the data in more detail.

Creating Cohorts

Using the CPS we construct a number of cohorts based on year of birth. The divisions we use are presented in Table C.1.1. The fist column refers to year born with the corresponding definition in the right column. Our results are robust to different definitions of the generation (for example taking smaller generation groups).

Table C.1: Different Birth Cohorts.

2000 – Present	Generation Z
1980 – 1999	Millennial's (Gen. Y)
1965 – 1979	Generation X
1946 – 1964	Baby Boomer's
1925 – 1945	Silent Generation

We then collapse of micro data to create aggregates by each generation, in order to look at trends in incomes for each of the generations. Our results are robust to how we classify the cohorts.

C.1.2 Summary of the Data

In Table C.2, summary statistics for the CPS are presented, both for each of our cohorts and for the total sample. Statistics are produced using the individual weights. All monetary amounts are adjusted for inflation using CPI with 1999 as the base year. We make a number of sample restrictions, firstly we drop individuals who are self-employed, in education or working for the government. And secondly we consider only individuals between the ages of 23 and 65.

Table C.2: Summary Statistics (CPS), Total and by Cohort

	Total		***Silent***		***Boomer's***		***Gen. X***		***Millenials***	
	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.	Mean	Std.Dev.
Demographics										
Age	40.41	11.27	44.26	10.58	40.52	11.04	35.13	7.19	28.27	3.84
Female	0.45	0.50	0.42	0.49	0.47	0.50	0.47	0.50	0.47	0.50
Married	0.68	0.47	0.78	0.41	0.68	0.47	0.62	0.49	0.42	0.49
Separated/Divorced	0.12	0.33	0.12	0.32	0.15	0.36	0.12	0.33	0.07	0.25
White	0.85	0.36	0.88	0.32	0.85	0.35	0.81	0.39	0.78	0.41
African American	0.10	0.30	0.09	0.29	0.09	0.29	0.10	0.30	0.11	0.32
Hispanic	0.21	0.41	0.28	0.45	0.12	0.32	0.19	0.39	0.23	0.42
Education										
High School Graduate	0.66	0.47	0.40	0.49	0.73	0.45	0.87	0.33	0.90	0.30
College	0.19	0.39	0.09	0.28	0.19	0.39	0.29	0.45	0.30	0.46
Labour Market										
Union	0.10	0.30	0.14	0.35	0.11	0.32	0.08	0.27	0.06	0.23
Hours Worked per Week	40.10	10.41	39.50	10.84	40.39	10.45	40.40	10.08	39.34	9.91
Hourly Wage	15.00	17.88	14.92	14.16	15.33	18.29	15.36	19.58	12.78	17.98
Labour Income	31,616	33,208	30,892	26,195	32,779	35,363	33,062	38,521	26,548	31,123
Total Income	33,751	34,821	33,454	28,113	35,017	37,163	34,692	39,808	27,684	31,835
Observations	2,724,724		603,676		1,095,058		594,059		207,060	

Notes: The sample used includes only individuals who are in employment, and are not self employed or working for the Government. We include those between the ages of 23 and 65.

We calculate the Hourly wage as the income from labour divided by the usual hours worked per week last year times by the 52 weeks of the year. We include a number of dummy variables, including *College* which is equal to one if the individual has at least a bachelors degree. A number of ethnicity dummies such as *Hispanic Origin* or *African American.* And lastly a dummy which is equal to one if the individual is a member of a trade union which is captured in the variable *Union.*

C.1.3 Economic Census

The Economic Census is available every five years since 1977. The survey contains data regarding seven primary industries; retail trade, wholesale trade, service industries, finance industries, construction industries, manufacturers and utilities and transport. The data coverage varies depending on the wave of the survey and by geographic level. Typically, we group Utilities and Transportation industries for our analysis.

There are a number of series which are available across all years and industries: Geographic Series, Non-Employer Statistics and Subject Series are the main available data series. We will predominately be using the data from the geographic series, which contains detailed informations about establishments which have payroll. Data is organised by kind of business and geographic areas; U.S., States, Metropolitan Areas, Counties and Places. The earlier data was aggregated mostly to broad SIC levels.

Where we have observations we can observe the number of establishments, annual payroll in $1000, value of first quarter payroll and value of sales and receipts. It is important to note that not all industries within a particular sector are covered by the economic census. These include; schools (all levels), U.S. postal service, public administration, private households and membership organisations.

Another feature of the service industry part of the economic census is that it is separated between those industries subject to federal income tax and those which are not.

In addition to matching at the broad state level and 2 digit industry we can also consider the MSA level both a 1-digit and 2-digit merge, this helps, to a certain extent address the issue of observation in the CPS who can only be matched to a broad sector rather than a 2-digit SIC code.

Creating Consistent Industry Codes

The CPS contains consistent industry codes from $1968 - present$ using the 1990 census classification. However, the Economic census makes use of SIC codes and then later NAICS coding, see Table C.3 for a summary of the industry codes which are used when. Firstly using NAICS crosswalks we create a file which maps each NAICS code to each other.[1] Following this we use the cross-walks of Autor et al. (2013) and Autor et al. (2017a) to create the consistent industry codes across SIC and match our datasets using a combination of SIC and NAICs codes. When we have constructed a consistent coding in the census data, we then match this to the CPS data at varying levels of disaggregation.

Table C.3: Industry Coding by Year in the Economic Census.

Year	**Industry Code**
1977-1982	1972 SIC Code
1987-1997	1987 SIC Code
1997-2012	Year Specific NAICS Code

Another issue is that not all levels of disaggregation in the industry codes are available every year. So as we increase the number of digit of disaggregation, what we gain in more finer detail we lose in terms of the time series. Most notably, 1977 and 1982 contain the most aggregating industry coding and as such as we lose these years of observations.

[1]See https://www.census.gov/eos/www/naics/concordances/concordances.html for NAICS crosswalks

Pre-1997 Data

Here, the industry coding used is not consistent and the available granularity of the data is year and industry dependant. We have at the two-digit industry code the sales and number of establishments, however we do not have number of employees or the payroll information. In some cases the best we can do is to aggregate at the industry level only.

Post-1997 Data

The raw economic census data provided breakdowns from two digit codes up to six digit codes for a range of industries. For the industries where the broadest classification was three digit, we summed across to collapse to a two-digit industry code which could be merged with the 1990 industry codes. We aggregated the flags such that, there is just an indicator as to whether a flag was present for one of the observations at the three digit level, but does not specify the type of special condition which the flag represented.

C.1.4 Merging the Data

We merge the CPS and Economic Census using a number of geographic and industry identifiers. We aggregate to various industry and geographic levels to merge with the industry level data. Notably, we can merge at the state and MSA level. Although MSA identifiers are not available at the two digit industry codes for year 1977 and 1982.

These are sufficient to merge with the available firm level microdata. In the CPS, around 4.4% of the observations have an unidentified state. Additionally, industries are identified using SIC codes in the CPS data. Additionally, we can include the most aggregate version of the data which will be at the national level and will be merged using the broad industry classifications.

As discussed, the Economic Census does not cover the universe of industries and occupations. As a result, for the purpose of merging the two data sets, we drop from our CPS sample those industries which are not covered by the Census. This equates to dropping just over 10% of the observations. The industries which are excluded are presented below in Table C.4.

Table C.4: Summary of Excluded Industries

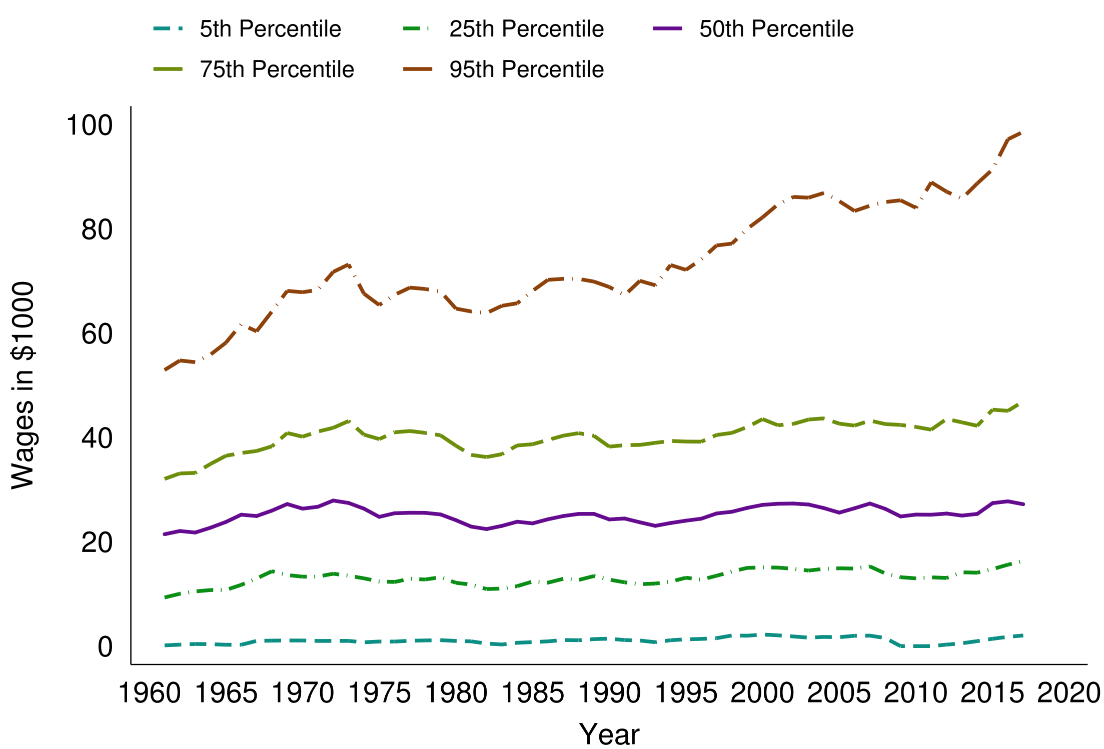
1990 Census Industry	**Title**
10-32	Agriculture, Forestry & Fisheries
400	Railroads
412	U.S. Postal Service
710	Security, commodity brokerage, and investment companies
711	Insurance
873	Labour Unions
880	Religious Organisations
881	Membership Organisations
¿900	Public Administration

We can then classify the industry codes into a broad 1-digit industry type. Which is the broadest level of the merging and matching. We are able to match SIC codes and industry census codes confidently at the 2 digit level. Then adding additional variation with merges at different geographical levels. Where there are gaps in the crosswalks, that is an 1990 industry code does not map to a 2-digit industry code we impute this ourselves if possible. For some we are not able to impute a 2 digit naics code for example, an 1990 industry code of *392 - Manufacturing, n.s (Not specified)* could refer to a naics sector code of either 31, 32 or 33.

We lose around 5% of observations each year during the merge. More so in the earlier years as the disaggregation in the 1977 and 1982 census was not as detailed as in the later census, as a result of the more limited data availability in the earlier years we drop the 1977 and 1982 waves of data.

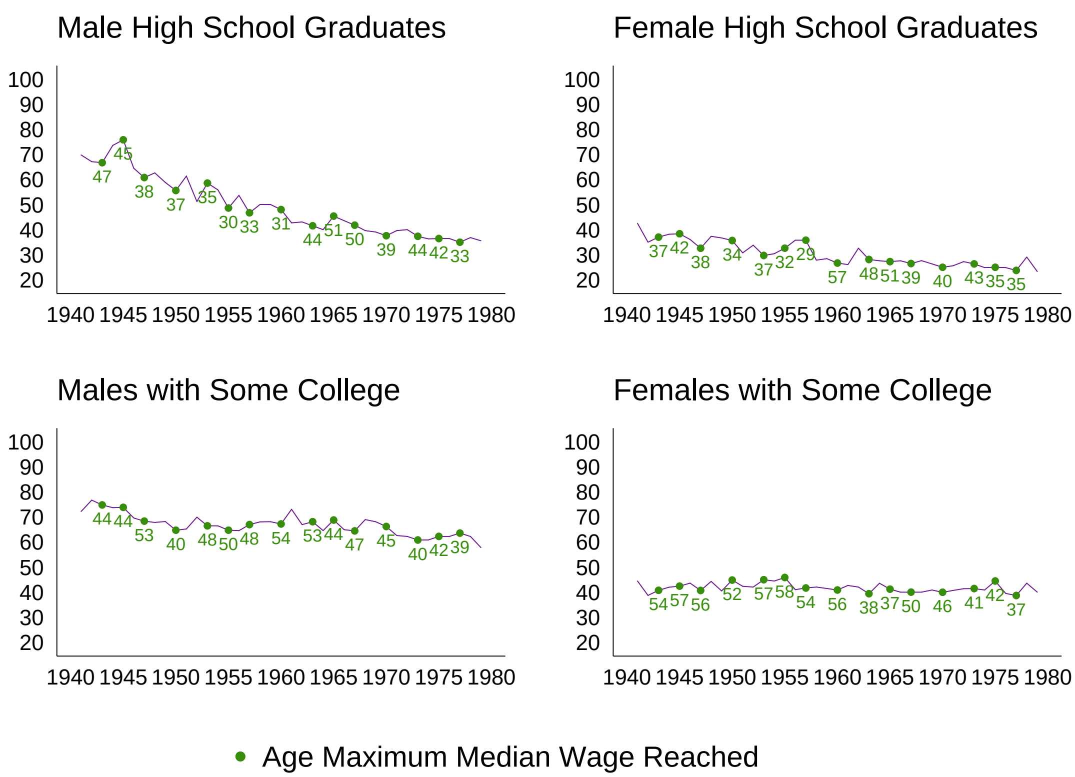
C.2 Additional Results

Figure C.1: Percentiles of the Earnings Distribution by Year



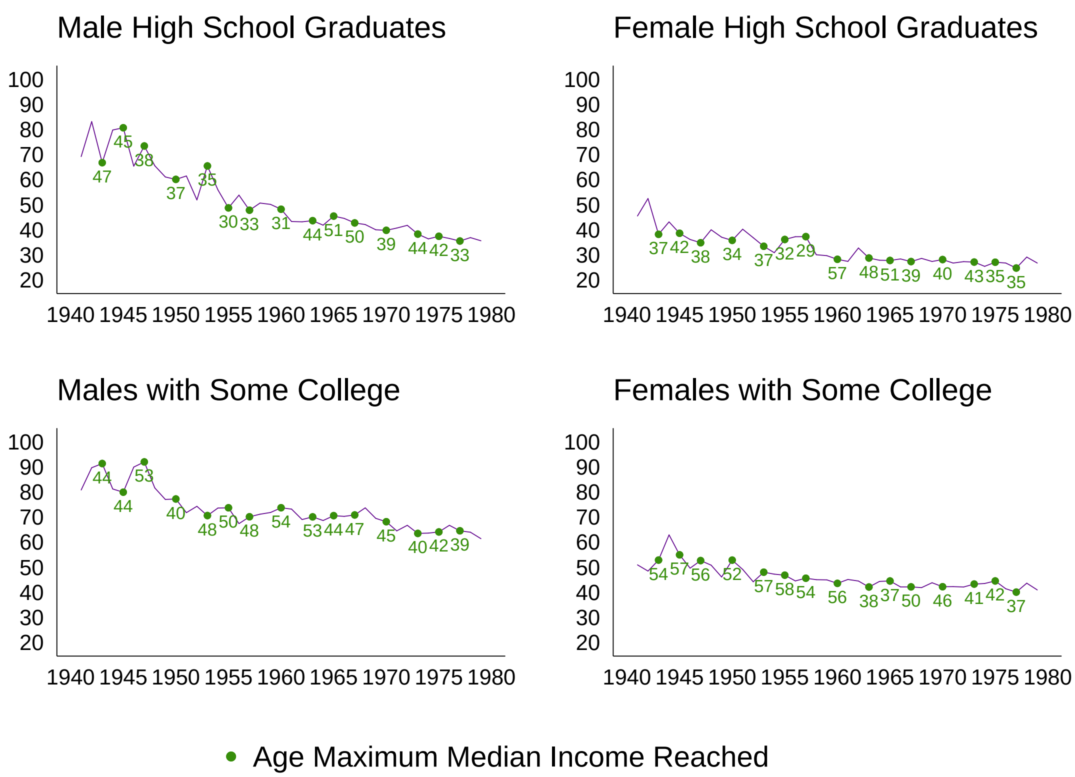
Source: Current Population Survey (CPS)

Figure C.2: Maximum Median Wage (in $1000) by Year born and the Age which it was reached.



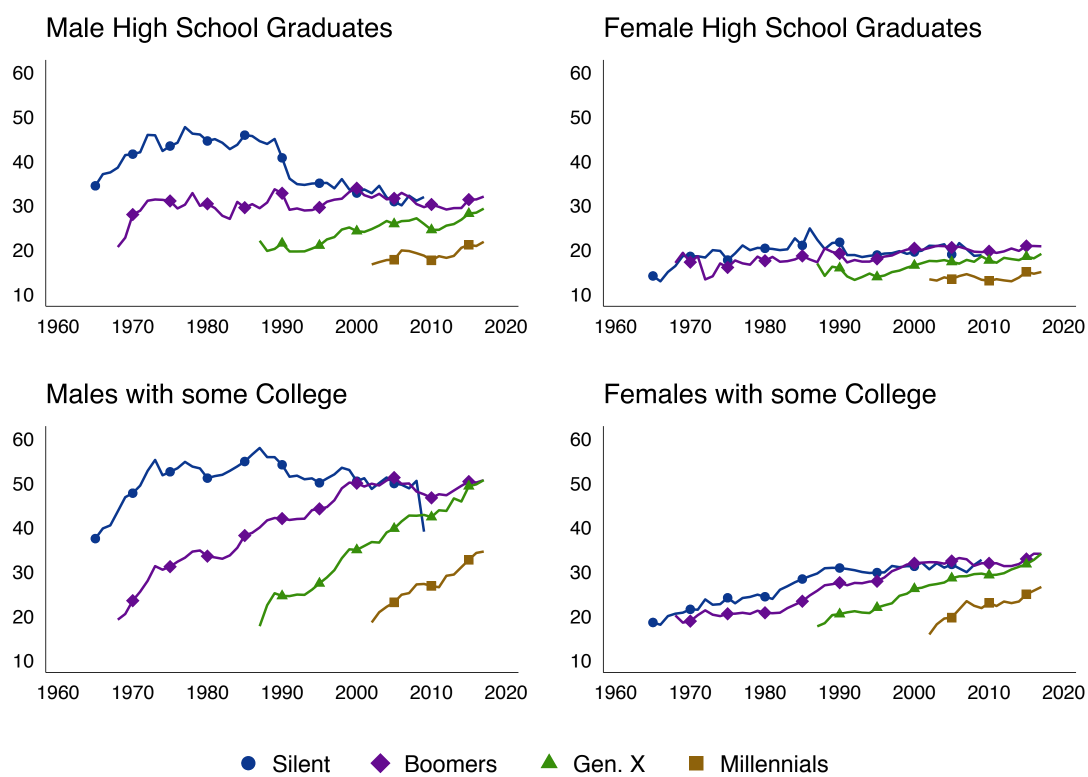
Source: Current Population Survey (CPS)

Figure C.3: Maximum Median Income (in $1000) by Year born and the Age which it was reached.



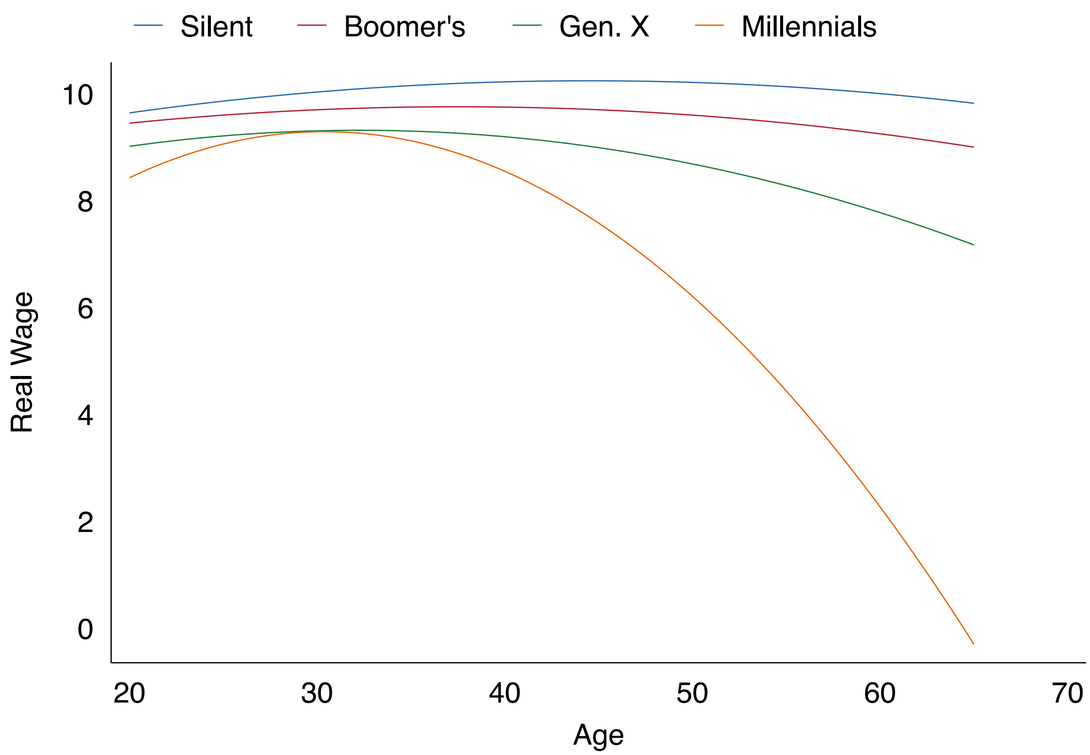
Source: Current Population Survey (CPS)

Figure C.4: Median Income (in $1000) for each Generation over time



Source: Current Population Survey (CPS)

Figure C.5: Predicted Median Income (in $1000) for each Generation by Age



Source: Each curve is computed given the coefficients on Age, Age^2 and the constant term in Table C.6

.

Table C.5: Regression on Wage by Generation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Silent	***Boomer's***	***Gen. X***	***Millenials***	***Silent***	***Boomer's***	***Gen. X***	***Millenials***
	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE
Age	0.072***	0.052***	0.152***	0.492***	0.063***	0.048***	0.147***	0.460***
	(0.005)	(0.002)	(0.004)	(0.032)	(0.005)	(0.002)	(0.004)	(0.032)
Age Sq	-0.001***	-0.001***	-0.002***	-0.008***	-0.001***	-0.000***	-0.002***	-0.007***
	(0.000)	(0.000)	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)	(0.001)
African American	-0.202***	-0.198***	-0.169***	-0.176***	-0.195***	-0.203***	-0.170***	-0.167***
	(0.012)	(0.007)	(0.009)	(0.018)	(0.012)	(0.007)	(0.009)	(0.018)
Hispanic	-0.243***	-0.201***	-0.133***	-0.054***	-0.236***	-0.196***	-0.131***	-0.059***
	(0.012)	(0.006)	(0.007)	(0.013)	(0.012)	(0.006)	(0.007)	(0.013)
High School Graduate	0.276***	0.274***	0.374***	0.358***	0.288***	0.280***	0.368***	0.350***
	(0.007)	(0.005)	(0.009)	(0.019)	(0.007)	(0.005)	(0.009)	(0.019)
College	0.396***	0.498***	0.556***	0.477***	0.428***	0.502***	0.541***	0.464***
	(0.012)	(0.005)	(0.006)	(0.012)	(0.012)	(0.005)	(0.006)	(0.012)
Female	-0.794***	-0.541***	-0.452***	-0.310***	-0.696***	-0.485***	-0.417***	-0.256***
	(0.007)	(0.004)	(0.005)	(0.011)	(0.007)	(0.004)	(0.006)	(0.011)
Retail Trade					0.000	0.000	0.000	0.000
					(.)	(.)	(.)	(.)
Wholesale Trade					0.378***	0.372***	0.343***	0.356***
					(0.016)	(0.010)	(0.014)	(0.029)
Services					0.146***	0.182***	0.196***	0.171***
					(0.012)	(0.007)	(0.008)	(0.014)
Finance					0.365***	0.397***	0.394***	0.394***
					(0.018)	(0.010)	(0.013)	(0.022)
Utilities & Transportation					0.494***	0.446***	0.345***	0.346***
					(0.014)	(0.008)	(0.011)	(0.021)
Manufacturing					0.451***	0.396***	0.334***	0.315***
					(0.010)	(0.006)	(0.009)	(0.018)
Construction					0.219***	0.211***	0.254***	0.316***
					(0.017)	(0.009)	(0.011)	(0.019)
Mining					0.559***	0.615***	0.563***	0.879***
					(0.025)	(0.014)	(0.025)	(0.041)
Constant	8.764***	8.901***	6.845***	1.976***	8.610***	8.685***	6.714***	2.269***
	(0.117)	(0.032)	(0.072)	(0.441)	(0.115)	(0.031)	(0.071)	(0.436)
Observations	66183	177964	94783	26604	66183	177964	94783	26604
R^2	0.234	0.201	0.227	0.163	0.265	0.227	0.245	0.190

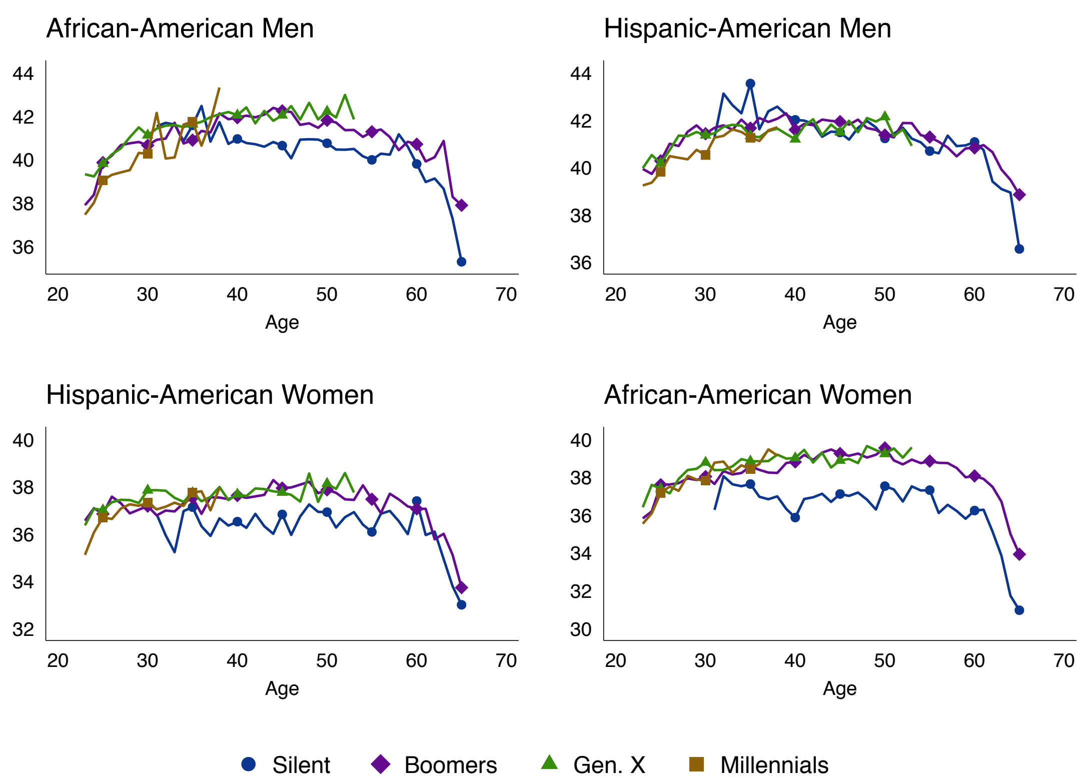
Notes: This table presents estimates of equation (4.3) but pooling across industries and disaggregating instead by industry, including covariates but not fixed effects. The dependent variable is log wages. The specification estimated is: $\log w^g_{j,t} = X'_{jt}\beta^g + \varepsilon^g_{j,t}$. $\varepsilon^g_{j,t}$ are clustered by state and by year. The data from the CPS and BEA are merged at the state geographic level and 1 digit industry code as described in Appendix C.1. The generation variables are all dummy variables defined on the basis of date of birth, as per Table 4.1. The omitted category is the Silent Generation. Standard Errors are in parenthesis. *** Significant at the 1 percent level.** Significant at the 5 percent level. * Significant at the 10 percent level.

Table C.6: Regression on Labour Share by Generation

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Silent	***Boomer's***	***Gen. X***	***Millenials***	***Silent***	***Boomer's***	***Gen. X***	***Millenials***
	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE	β / SE
Age	0.012***	0.005***	0.023***	0.014***	0.005***	0.001***	0.020***	0.011***
	(0.000)	(0.000)	(0.001)	(0.003)	(0.000)	(0.000)	(0.001)	(0.002)
Age Sq	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
African American	0.002**	0.011***	0.017***	0.007***	-0.006***	-0.001	0.003**	0.001
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Hispanic	-0.006***	-0.006***	0.001	0.004***	-0.010***	-0.011***	0.001	-0.000
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
High School Graduate	-0.026***	-0.004***	0.026***	0.017***	-0.025***	-0.022***	0.001	0.001*
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
College	0.005***	0.026***	0.030***	0.023***	-0.005***	0.002**	0.005***	0.006***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Female	-0.002***	0.021***	0.021***	0.012***	-0.004***	0.000	-0.000	-0.002***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.001)	(0.001)
Retail Trade					0.000	0.000	0.000	0.000
					(.)	(.)	(.)	(.)
Wholesale Trade					-0.021***	-0.020***	-0.021***	-0.022***
					(0.001)	(0.000)	(0.000)	(0.000)
Services					0.077***	0.149***	0.115***	0.063***
					(0.000)	(0.000)	(0.000)	(0.000)
Finance					0.083***	0.422***	0.446***	0.253***
					(0.003)	(0.004)	(0.004)	(0.004)
Utilities & Transportation					0.038***	0.284***	0.285***	0.120***
					(0.001)	(0.004)	(0.004)	(0.002)
Manufacturing					0.097***	0.086***	0.009***	-0.011***
					(0.001)	(0.000)	(0.000)	(0.000)
Construction					0.036***	0.057***	0.045***	0.016***
					(0.000)	(0.000)	(0.000)	(0.000)
Mining					0.023***	0.040***	-0.002**	-0.003***
					(0.002)	(0.001)	(0.001)	(0.001)
Constant	-0.107***	0.064***	-0.340***	-0.214***	0.008	0.076***	-0.335***	-0.182***
	(0.009)	(0.005)	(0.012)	(0.038)	(0.006)	(0.003)	(0.009)	(0.025)
Observations	51266	162253	97398	28045	51266	162253	97398	28045
R^2	0.218	0.014	0.037	0.070	0.585	0.353	0.492	0.651

Notes: This table presents estimates of equation (4.3) but pooling across industries and disaggregating instead by industry, including covariates but not fixed effects. The dependent variable is the labour share. The specification estimated is: $\lambda^g_{j,t} = X'_{jt}\beta^g + \delta_t + \delta_s + \varepsilon^g_{j,t}$. $\varepsilon^g_{j,t}$ are clustered by state and by year. The data from the CPS and BEA are merged at the state geographic level and 1 digit industry code as described in Appendix C.1. The generation variables are all dummy variables defined on the basis of date of birth, as per Table 4.1. The omitted category is the Silent Generation. Standard Errors are in parenthesis. *** Significant at the 1 percent level.** Significant at the 5 percent level. * Significant at the 10 percent level.

Figure C.6: Hours Worked by Generation over the Lifecycle



Source: Current Population Survey (CPS)

Bibliography

Abowd, J. M., F. Kramarz, and D. N. Margolis (1999): "High wage workers and high wage firms," *Econometrica*, 67, 251–333.

Acemoglu, D. and D. Autor (2011): *Skills, tasks and technologies: Implications for employment and earnings*, vol. 4, Elsevier Inc.

Acemoglu, D. and P. Restrepo (2018): "The race between man and machine: Implications of technology for growth, factor shares, and employment," .

Almås, I., A. W. Cappelen, J. T. Lind, E. Sørensen, and B. Tungodden (2011a): "Measuring unfair (in)equality," *Journal of Public Economics*, 95, 488–499.

Almås, I., T. Havnes, and M. Mogstad (2011b): "Baby booming inequality? Demographic change and earnings inequality in Norway, 1967-2000," *Journal of Economic Inequality*, 9, 629–650.

Alvaredo, F., A. B. Atkinson, L. Chancel, T. Piketty, E. Saez, G. Zucman, et al. (2016): "Distributional National Accounts (DINA) Guidelines: Concepts and Methods used in WID. world," *WID.World Working Paper*, 2.

Andrews, D. W., W. Kim, and X. Shi (2017): "Commands for testing conditional moment inequalities and equalities," *The Stata Journal*, 17, 56–72.

Andrews, D. W. and X. Shi (2013): "Inference based on conditional moment inequalities," *Econometrica*, 81, 609–666.

Andrews, D. W. and G. Soares (2010): "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 78, 119–157.

Atack, J. and F. Bateman (1979): "The measurement and trend of inequality," *Economics Letters*, 4, 389–393.

Atkinson, A. B. (1971): "The distribution of wealth and the individual lifecycle," *Oxford Economic Papers*, 23, 239–254.

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

Autor, D., D. Dorn, and G. H. Hanson (2013): "The China syndrome: Local labor market effects of import competition in the United States," *American Economic Review*, 103, 2121–68.

——— (2017a): "When Work Disappears: Manufacturing Decline and the Falling Marriage-Market Value of Young Men," .

Autor, D., D. Dorn, L. Katz, C. Patterson, and J. Van Reenen (2017b): "The Fall of the Labor Share and the Rise of Superstar Firms," *Working Paper*.

Autor, D. and A. Salomons (2018): "Is Automation Labor Share-Displacing? Productivity Growth, Employment, and the Labor Share," *Brookings Papers on Economic Activity*, 2018, 1–87.

Bengtsson, E., A. Missiaia, M. Olsson, and P. Svensson (2017): "Wealth inequality in Sweden 1750–1900," *The Economic History Review*.

Black, S. and P. Devereux (2011): "Recent Developments in Intergenerational Mobility," *Handbook of Labor Economics*, 4, 1487–1541.

Blanden, J., P. Gregg, and L. Macmillan (2007): "Accounting for intergenerational income persistence: noncognitive skills, ability and education," *The Economic Journal*, 117, C43–C60.

Bloom, N., R. Sadun, and J. Van Reenen (2012): "Americans do IT better: US multinationals and the productivity miracle," .

Blundell, R., A. Bozio, and G. Laroque (2011): "Labor supply and the extensive margin," *American Economic Review*, 101, 482–486.

Blundell, R. and B. Etheridge (2010): "Consumption, income and earnings inequality in Britain," *Review of Economic Dynamics*.

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

Blundell, R. and I. Preston (1998): "Consumption inequality and income uncertainty," *The Quarterly Journal of Economics*.

Blundell, R., H. Reed, and T. M. Stoker (2003): "Interpreting aggregate wage growth: The role of labor market participation," *American Economic Review*, 93, 1114–1131.

Borella, M., M. De, N. F. Yang, J. Altonji, R. Blundell, M. C. Dias, Z. Eckstein, R. Lenz, D. Neal, J. Skinner, M. De Nardi, and F. Yang (2017): "The Effects of Marriage-Related Taxes and Social Security Benefits," .

BORELLA, M., M. D. NARDI, AND F. YANG (2019): "The Lost Ones : the Opportunities and Outcomes of Non-College Educated Americans Born in the 1960s," *NBER Macroeconomics Annual*.

BOURGUIGNON, F., F. H. FERREIRA, AND P. G. LEITE (2008): "Beyond Oaxaca-Blinder: Accounting for differences in household income distributions," *Journal of Economic Inequality*, 6, 117–148.

BOWLES, S. AND H. GINTIS (2002): "The inheritance of inequality," *Journal of Economic Perspectives*, 16, 3–30.

BREWER, M. AND L. WREN-LEWIS (2016): "Accounting for Changes in Income Inequality: Decomposition Analyses for the UK, 1978-2008," *Oxford Bulletin of Economics and Statistics*, 78, 289–322.

BURSTEIN, A., E. MORALES, AND J. VOGEL (2019): "Changes in Between-Group Inequality: Computers, Occupations, and International Trade," *American Economic Journal: Macroeconomics*, 11, 348–400.

CAGETTI, M. AND M. DE NARDI (2006): "Entrepreneurship, Frictions, and Wealth," *Journal of Political Economy*, 114, 835–870.

CANAY, I. A. AND A. M. SHAIKH (2017): "Practical and theoretical advances in inference for partially identified models," in *Advances in Economics and Econometrics: Eleventh World Congress*, Cambridge University Press Cambridge, vol. 2, 271–306.

CASE, A. AND A. DEATON (2015): "Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century," *Proceedings of the National Academy of Sciences*, 112, 15078–15083.

CHADWICK, L. AND G. SOLON (2002): "Intergenerational income mobility among daughters," *American Economic Review*, 92, 335–344.

CHERNOZHUKOV, V., S. LEE, AND A. M. ROSEN (2013): "Intersection bounds: estimation and inference," *Econometrica*, 81, 667–737.

CHETTY, R., D. GRUSKY, M. HELL, N. HENDREN, R. MANDUCA, AND J. NARANG (2017): "The fading American dream: Trends in absolute income mobility since 1940," *Science*, 356, 398–406.

CHETTY, R., N. HENDREN, P. KLINE, AND E. SAEZ (2014a): "Where is the land of opportunity? The geography of intergenerational mobility in the United States," *The Quarterly Journal of Economics*, 129, 1553–1623.

CHETTY, R., N. HENDREN, P. KLINE, E. SAEZ, AND N. TURNER (2014b): "Is the United States still a land of opportunity? Recent trends in intergenerational mobility," *American Economic Review*, 104, 141–147.

Chetty, R., N. Hendren, F. Lin, J. Majerovitz, and B. Scuderi (2016): "Gender Gaps in Childhood: Skills, Behavior, and Labor Market Preparedness Childhood Environment and Gender Gaps in Adulthood," *American Economic Review*, 106, 282–288.

Corneo, G. (2015): "Income inequality from a lifetime perspective," *Empirica*, 42, 225–239.

Couch, K. A. and D. R. Lillard (1998): "Sample selection rules and the intergenerational correlation of earnings," *Labour Economics*, 5, 313–329.

Cowell, F. A. and S. P. Jenkins (1995): "How Much Inequality Can we Explain ? A Methodology and an Application to the United States," *Economic Journal*, 105, 421–430.

Creedy, J., N. Gemmell, and L. Nguyen (2017): "Income inequality in New Zealand, 1935-2014," Tech. rep.

De Haan, M. (2011): "The effect of parent's schooling on child's schooling: A nonparametric bounds analysis," *Journal of Labor Economics*, 29, 859–892.

Elsby, M., B. Hobijn, and A. Sahin (2013): "The Decline of the U.S. Labor Share," *Brookings Papers on Economic Activity*, 44, 1–63.

Fernández, R., A. Fogli, and C. Olivetti (2004): "Mothers and sons: Preference formation and female labor force dynamics," *The Quarterly Journal of Economics*, 119, 1249–1299.

Formby, J., T. Seaks, and W. Smith (1989): "On the measurement and trend of inequality: a reconsideration," *American Economic Review*, 79, 256–264.

Formby, J. P. and T. G. Seaks (1980): "Paglin's Gini Measure of Inequality: A Modification," *American Economic Review*, 70, 479–482.

Friedman, B. M. (2005): *The Moral Consequences of Economic Growth*, New York: Alfred A. Knopf.

Fryer, R. (2011): "Racial inequality in the 21st century: The declining significance of discrimination," in *Handbook of Labor Economics*, Elsevier, vol. 4, 855–971.

Fuchs-Schündeln, N., D. Krueger, and M. Sommer (2010): "Inequality trends for Germany in the last two decades: A tale of two countries," *Review of Economic Dynamics*.

Gibbons, S., H. G. Overman, and P. Pelkonen (2014): "Area disparities in Britain: Understanding the contribution of people vs. place through variance decompositions," *Oxford Bulletin of Economics and Statistics*, 76, 745–763.

Goldin, C. (2014): "A Grand Gender Convergence: Its Last Chapter," *American Economic Review*, 104, 1091–1119.

Gordon, R. J. (2017): *The rise and fall of American growth: The US standard of living since the civil war*, vol. 70, Princeton University Press.

Gorr, M. (1983): "Rawls on Natural Inequality," *The Philosophical Quarterly*, 33, 1.

Grossman, G., E. Helpman, E. Oberfield, and T. Sampson (2018): "The Productivity Slowdown and the Declining Labor Share," 2018 Meeting Papers 169, Society for Economic Dynamics.

Guvenen, F., G. Kaplan, J. Song, and J. Weidner (2017): "Lifetime Incomes in the United States Over Six Decades," .

Heathcote, J., F. Perri, and G. L. Violante (2010): "Unequal we stand: An empirical analysis of economic inequality in the United States, 1967–2006," *Review of Economic Dynamics*, 13, 15–51.

Heckman, J. J. (1979): "Sample Selection Bias as a Specification Error," *Econometrica*, 47, 153–161.

Ho, K. and A. M. Rosen (2017): "Partial Identification in Applied Research: Benefits and Challenges," in *Advances in Economics and Econometrics: Volume 2: Eleventh World Congress*, Cambridge University Press, vol. 2, 307.

Huggett, M., G. Ventura, and A. Yaron (2006): "Human capital and earnings distribution dynamics," *Journal of Monetary Economics*, 53, 265–290.

——— (2011): "Sources of lifetime inequality," *American Economic Review*, 101, 2923–2954.

Human Fertility Database (2013): "Human Fertility Database," Available at www.humanfertility.org (data downloaded July 2013). Max Planck Institute for Demographic Research (Germany) and Vienna Institute of Demography (Austria).

Human Mortality Database (2013): "Human Mortality Database." University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany) www.mortality.org or www.humanmortality.de (data downloaded July 2013).

Imbens, G. W. and C. F. Manski (2004): "Confidence intervals for partially identified parameters," *Econometrica*, 72, 1845–1857.

Jappelli, T. and L. Pistaferri (2010): "Does consumption inequality track income inequality in Italy?" *Review of Economic Dynamics*.

Jones, C. I. and P. J. Klenow (2016): "Beyond GDP? Welfare across countries and time," *American Economic Review*, 106.

Jorgenson, D. W., M. S. Ho, and K. J. Stiroh (2008): "A retrospective look at the U.S. productivity growth resurgence," in *Journal of Economic Perspectives*.

Juhn, C. and S. Potter (2006): "Changes in labor force participation in the United States," *Journal of Economic Perspectives*, 20, 27–46.

Kaplan, G. and S. Schulhofer-Wohl (2018): "The Changing (Dis-)utility of Work," *Journal of Economic Perspectives*, 32, 239–58.

Karabarbounis, L. and B. Neiman (2014): "The Global Decline of the Labor Share," *The Quarterly Journal of Economics*, 129, 61–103.

King, R. G., C. I. Plosser, and S. T. Rebelo (1988): "Production, growth and business cycles: I. The basic neoclassical model," *Journal of Monetary Economics*, 21, 195–232.

Lambert, P. J. and J. R. Aronson (1993): "Inequality decomposition analysis and the Gini coefficient revisited," *Economic Journal*, 1221–1227.

Lang, K. and M. Manove (2011): "Education and Labor Market Discrimination," *American Economic Review*, 101, 1467–1496.

Lee, C.-I. and G. Solon (2009): "Trends in intergenerational income mobility," *The Review of Economics and Statistics*, 91, 766–772.

Lee, S., K. Song, and Y.-J. Whang (2013): "Testing functional inequalities," *Journal of Econometrics*, 172, 14–32.

Lerman, R. I. . and S. Yitzhaki (1985): "Income Inequality Effects by Income Source : A New Approach and Applications to the United States," *The Review of Economics and Statistics*, 67, 151–156.

Leslie, P. H. (1945): "The use of matrices in certain population mathematics," *Biometrika*, 33, 183–212.

——— (1948): "Some Further Notes on the Use of Matrices in Population Mathematics," *Biometrika*, 35, 213.

Mankiw, N. G. and M. Weinzierl (2006): "Dynamic scoring: a back-of-the-envelope guide," *Journal of Public Economics*, 90, 1415–1433.

Manski, C. F. (1989): "Anatomy of the Selection Problem," *Journal of Human Resources*, 343–360.

——— (1994): "The selection problem," in *Advances in Econometrics, Sixth World Congress*, ed. by C. Sims, Cambridge, U.K.: Cambridge University Press, vol. 1, 143–70.

Manski, C. F. and J. V. Pepper (2000): "Monotone instrumental variables: With an application to the returns to schooling," *Econometrica*, 68, 997–1010.

——— (2009): "More on monotone instrumental variables," *The Econometrics Journal*, 12, S200–S216.

Mazumder, B. and D. I. Levine (2002): "Choosing the right parents: Changes in the intergenerational transmission of inequality between 1980 and the early 1990s," .

McDaniel, C. (2011): "Forces shaping hours worked in the OECD, 1960-2004," *American Economic Journal: Macroeconomics*, 3, 27–52.

Milanovic, B. (1997): "A simple way to calculate the Gini coefficient, and some implications," *Economics Letters*, 56, 45–49.

——— (2016): "Income inequality is cyclical," *Nature*, 537, 479–482.

Minicozzi, A. L. (2003): "Estimation of sons' intergenerational earnings mobility in the presence of censoring," *Journal of Applied Econometrics*, 18, 291–314.

Mookherjee, D. and A. Shorrocks (1982): "A decomposition analysis of the trend in UK income inequality," *Economic Journal*, 886–902.

Neal, D. and S. Rosen (2000): "Theories of the distribution of earnings," *Handbook of Income Distribution*, 1, 379–427.

Paglin, M. (1975): "The measurement and trend of inequality: a basic revision," *American Economic Review*, 598–609.

Pesaran, M. H. and R. Smith (1995): "Estimating long-run relationships from dynamic heterogeneous panels," *Journal of Econometrics*, 68, 79–113.

Piketty, T. (2003): "Income Inequality in France, 1901–1998," *Journal of Political Economy*, 111, 1004–1042.

——— (2011): "On the long-run evolution of inheritance France 1820-2050," *The Quarterly Journal of Economics*, 126, 1071–1131.

Piketty, T. and E. Saez (2003a): "Income Inequality in the United States, 1913–1998," *The Quarterly Journal of Economics*, 118, 1–39.

——— (2003b): "Income inequality in the United States, 1913–1998," *The Quarterly Journal of Economics*, 118, 1–41.

——— (2014): "Inequality in the long run," *Science*, 344.

Roine, J. and D. Waldenström (2015): "Long-Run Trends in the Distribution of Income and Wealth," in *Handbook of Income Distribution*, vol. 2, 469–592.

Saez, E. and G. Zucman (2016): "Wealth Inequality in the United States since 1913: Evidence from Capitalized Income Tax Data," *Quarterly Journal of Economics*, 131, 519–578.

Song, J., D. J. Price, N. Bloom, F. Guvenen, and T. von Wachter (2016): "Firming up inequality," .

Stoye, J. (2009): "More on confidence intervals for partially identified parameters," *Econometrica*, 77, 1299–1315.

Toledo, M. (2007): "On the intergenerational persistence of work hours," .

Yamada, T. (2012): "Income risk, macroeconomic and demographic change, and economic inequality in Japan," *Journal of Economic Dynamics and Control*.