**Youth in Transition: Longitudinal Comparisons of Youth Transitions in the UK using Cohort and Synthetic Cohort Data**

Scott Oatley

**A black and white logo

Description automatically generated**

Thesis submitted for the degree of Doctor of Philosophy

School of Social and Political Science

University of Edinburgh

2024

**Thesis Declaration**

I declare that this thesis has been composed by myself, is my own work, and has not been submitted, in whole or in part, for any other degree.

Scott Oatley

**Lay Summary**

This thesis aims to understand how structural factors, such as: education, sex, housing tenure, and social class affect young people’s transition from school-to-work. Prior research has established a strong relationship between structural inequalities and individual labour market outcomes. This previous research has primarily only focused on older datasets that do not allow the investigation of contemporary youth transitions. This thesis will apply contemporary statistical techniques to update this previous literature whilst also looking at contemporary datasets to investigate youth transitions across a larger timespan. This thesis analyses data from multiple different time points, using cohort studies and where appropriate creating groups called synthetic cohorts from longitudinal household panel data to study changes over time. This thesis examines individuals born between 1958 and 2000 to understand how changes over time affect school-to-work transitions.

This thesis finds that factors like: education, sex, housing tenure, and social class significantly affect how young people transition from school-to-work. The impact of these factors varies depending on the cohort studied, demonstrating that socio-historical context matters within the study of youth transitions. Education’s impact on youth transitions is strong but has a decreasing impact throughout each successive cohort studied. Social class-based affects remain rigid in the face of cohort change. Whilst sex and housing tenure-based effects change with each successive cohort.

**Abstract**

This thesis investigates how structural inequalities, such as education, sex, housing tenure, and social class, shape youth’s school-to-work transitions across different cohorts and different socio-historical contexts. Prior research has established a link between structural inequalities such as social class, sex, housing tenure, and educational attainment and their influence upon youth transitions. This thesis makes two key contributions: it modernises the analysis of youth transitions by incorporating contemporary statistical methods, including sensitivity analyses and advanced techniques for handling missing data. Additionally, it addresses a gap in the literature by examining transitions of individuals born in the 1980s and 1990s. Periods that have been underexplored in previous studies of the sociology of youth. Synthetic cohorts are constructed from two longitudinal household panel studies to bridge this gap, allowing for a longitudinal analysis of youth transitions over time. The analysis distinguishes between ‘First Transitions’ – whether individuals continued schooling after the end of the mandatory period – and ‘First Destinations – the different paths taken after leaving mandatory schooling. Focusing on these two early time points within a young person’s transition into adulthood provides a comprehensive examination of the life course at a pivotal stage in a young person’s life.

The thesis is structured into four parts: an introduction outlining the research context and theoretical framework, a detailed analysis of the first transition from schooling, an examination of post-school destinations, and concluding remarks that synthesis the findings across the thesis.

**Acknowledgements**

This PhD research was supported by an ERSC studentship. I am grateful for the support provided.

Contents

[List of Tables 10](#_Toc174390608)

[List of Appendix Tables 12](#_Toc174390609)

[List of Figures 13](#_Toc174390610)

[List of Abbreviations 16](#_Toc174390611)

[Part 1 Introduction 18](#_Toc174390612)

[Chapter 1.1 School-to-work transitions in context 20](#_Toc174390613)

[Chapter 1.2 Social Theory 22](#_Toc174390614)

[Chapter 1.3 Social Stratification 36](#_Toc174390615)

[1.3.1 Changes in the social stratification structures 37](#_Toc174390616)

[Chapter 1.4 The British Education System 38](#_Toc174390617)

[Chapter 1.5 The British (Youth) Labour Market 41](#_Toc174390618)

[Chapter 1.6 Research Questions 45](#_Toc174390619)

[Chapter 1.7 Data 47](#_Toc174390620)

[1.7.1 National Childhood Development Study 48](#_Toc174390621)

[1.7.1.1 Sample Size and Attrition 50](#_Toc174390622)

[1.7.2 British Cohort Survey 52](#_Toc174390623)

[1.7.2.1 Sample Attrition and missingness: 54](#_Toc174390624)

[1.7.3 The British Household Panel Survey 56](#_Toc174390625)

[1.7.4 United Kingdom Household Panel Survey (Understanding Society) 58](#_Toc174390626)

[1.7.4.1 Complex Survey Design 59](#_Toc174390627)

[Chapter 1.8 Methods 60](#_Toc174390628)

[1.8.1 Logistic Regression Models 61](#_Toc174390629)

[1.8.2 Multinominal Logistic Regression Models 61](#_Toc174390630)

[1.8.3 Goodness-of-fit statistics 62](#_Toc174390631)

[Chapter 1.9 Structure of Thesis 69](#_Toc174390632)

[Part 2 Youths First Major Transition Post-Mandatory Schooling 70](#_Toc174390633)

[Chapter 2.1 Introduction 70](#_Toc174390634)

[Chapter 2.2 Literature Review: Cohorts in Context 72](#_Toc174390635)

[2.2.1 NCDS in Context 73](#_Toc174390636)

[2.2.1.1 Story of transitions for NCDS youth 74](#_Toc174390637)

[2.2.1.2 Structural Barriers to successful transitions – the role of sex and social-class 86](#_Toc174390638)

[2.2.1.2.1 Sex 87](#_Toc174390639)

[2.2.1.2.2 Social Class 88](#_Toc174390640)

[2.2.1.2.3 Educational Attainment and training 88](#_Toc174390641)

[2.2.2 BCS in Context 90](#_Toc174390642)

[2.2.2.1 Story of transitions for BCS youth 91](#_Toc174390643)

[2.2.2.2 Structural Barriers to successful transitions – the role of social class and sex 101](#_Toc174390644)

[2.2.2.2.1 Social Class 101](#_Toc174390645)

[2.2.2.2.2 Sex 102](#_Toc174390646)

[2.2.2.2.3 Conclusion 102](#_Toc174390647)

[2.2.3 UKHLS in Context 103](#_Toc174390648)

[Chapter 2.3 Breakdown of Dataset Context 106](#_Toc174390649)

[2.3.1 NCDS Dataset context and methodology 106](#_Toc174390650)

[2.3.2 BCS Dataset context and methodology 107](#_Toc174390651)

[2.3.3 UKHLS Dataset context and methodology 108](#_Toc174390652)

[2.3.4 Introduction to Measures for Subsequent Analysis 110](#_Toc174390653)

[2.3.4.1 Economic Activity 110](#_Toc174390654)

[2.3.4.2 Educational Attainment 116](#_Toc174390655)

[2.3.4.3 Sex 120](#_Toc174390656)

[2.3.4.4 Race 121](#_Toc174390657)

[2.3.4.5 Housing Tenure 122](#_Toc174390658)

[2.3.4.6 Social Stratification Measures 124](#_Toc174390659)

[2.3.4.6.1 SOC Codes 130](#_Toc174390660)

[2.3.4.6.2 Registrar General Class Schema 136](#_Toc174390661)

[2.3.4.6.3 National Statistics Socio-Economic Classification 139](#_Toc174390662)

[2.3.4.6.4 CAMSIS 142](#_Toc174390663)

[2.3.4.7 Descriptive Statistics 145](#_Toc174390664)

[Chapter 2.4 Modelling First Major Transition 147](#_Toc174390665)

[2.4.1 Discussion and Conclusions 161](#_Toc174390666)

[Chapter 2.5 In-depth NCDS Analysis 166](#_Toc174390667)

[2.5.1 Descriptive Statistics 167](#_Toc174390668)

[2.5.2 Initial Model 179](#_Toc174390669)

[2.5.3 Sensitivity Analysis of Social Stratification Measures using NCDS 193](#_Toc174390670)

[2.5.3.1 Discussion and Conclusion 208](#_Toc174390671)

[2.5.4 SOC Code Sensitivity analysis using NCDS 211](#_Toc174390672)

[2.5.4.1 Discussion and Conclusions 242](#_Toc174390673)

[2.5.5 Handling Missing Data 242](#_Toc174390674)

[2.5.5.1 Missing Data 242](#_Toc174390675)

[2.5.5.2 Simulation of Handling Missing Data Strategies 254](#_Toc174390676)

[2.5.5.3 Handling Missing Data in the NCDS 261](#_Toc174390677)

[2.5.5.4 Discussion and Conclusions 272](#_Toc174390678)

[2.5.6 Discussion and Conclusions for Granular NCDS Analysis 273](#_Toc174390679)

[Chapter 2.6 Granular BCS Analysis 277](#_Toc174390680)

[2.6.1 Descriptive Statistics 278](#_Toc174390681)

[2.6.2 Initial Model 289](#_Toc174390682)

[2.6.3 Sensitivity Analysis of Social Stratification Measures using BCS 301](#_Toc174390683)

[2.6.3.1 Discussion and Conclusion 315](#_Toc174390684)

[2.6.4 SOC Code Sensitivity analysis using BCS 315](#_Toc174390685)

[2.6.4.1 Measuring SOC Codes 318](#_Toc174390686)

[2.6.4.2 Discussion and Conclusions 346](#_Toc174390687)

[2.6.5 Handling Missing Data in the BCS 346](#_Toc174390688)

[2.6.5.1 Discussion and Conclusions 357](#_Toc174390689)

[2.6.6 Discussion and Conclusions 358](#_Toc174390690)

[Chapter 2.7 Granular UKHLS Analysis 359](#_Toc174390691)

[2.7.1 Sensitivity Analysis of Social Stratification Measures using UKHLS 359](#_Toc174390692)

[2.7.1.1 Testing Measures of Parental Social Class 359](#_Toc174390693)

[2.7.1.2 Discussion and Conclusion 359](#_Toc174390694)

[2.7.2 SOC Code Sensitivity analysis using UKHLS 359](#_Toc174390695)

[2.7.2.1 Measuring SOC Codes 359](#_Toc174390696)

[2.7.2.2 Discussion and Conclusions 359](#_Toc174390697)

[2.7.3 Handling Missing Data in the UKHLS 359](#_Toc174390698)

[2.7.3.1 Discussion and Conclusions 359](#_Toc174390699)

[2.7.4 Discussion and Conclusions 359](#_Toc174390700)

[Chapter 2.8 A Return to Modelling First Transition 359](#_Toc174390701)

[2.8.1 Discussion and Conclusions 370](#_Toc174390702)

[Chapter 2.9 Discussion and Conclusions for Part 1 372](#_Toc174390703)

[Part 3 Youth’s First Destination post-mandatory schooling 378](#_Toc174390704)

[Chapter 3.1 First Destination 378](#_Toc174390705)

[3.1.1 Discussion and Conclusions 391](#_Toc174390706)

[Chapter 3.2 Discussion and Conclusions for Part 2 391](#_Toc174390707)

[Part 4 Conclusions 392](#_Toc174390708)

[Chapter 4.1 Introduction to Part 5 392](#_Toc174390709)

[Chapter 4.2 Substantive Conclusions 392](#_Toc174390710)

[Chapter 4.3 Methodological Reflections 392](#_Toc174390711)

[Chapter 4.4 Final Remarks 392](#_Toc174390712)

[Part 5 Appendix 393](#_Toc174390713)

[Part 6 Bibliography 405](#_Toc174390714)

###### of Tables

[Table 0.1 Sweeps Included in Analysis NCDS 50](#_Toc174390382)

[Table 0.2 Participation in the NCDS from birth to 23 years 50](#_Toc174390383)

[Table 0.3 Sweeps Included for Analysis BCS 52](#_Toc174390384)

[Table 0.4 Participation in the BCS from Birth to 30 years 55](#_Toc174390385)

[Table 2.1 Frequency Statistics for Economic Activity 111](#_Toc174390386)

[Table 2.2 Frequency Statistics for Economic Activity 113](#_Toc174390387)

[Table 2.3 Frequency Statistics for Economic Activity Part 2 114](#_Toc174390388)

[Table 2.4 Breakdown of classification of SOC 90 and SOC 2000 132](#_Toc174390389)

[Table 2.5 Sub-major groups of SOC 90 and SOC 2000 by Skill Level 132](#_Toc174390390)

[Table 2.6 RGSC Class Schema 138](#_Toc174390391)

[Table 2.7 NS-SEC Class Schema 141](#_Toc174390392)

[Table 2.8 Examples of Occupations from Analytical NS-SEC 142](#_Toc174390393)

[Table 2.9 Examples of CAMSIS scores by SOC-90 Codes 144](#_Toc174390394)

[Table 2.10 Pooled Cohort Descriptive Statistics 145](#_Toc174390395)

[Table 2.11 Pooled descriptive statistics by Cohort 146](#_Toc174390396)

[Table 2.12 Modelling First Major Transition with Combined Cohorts 150](#_Toc174390397)

[Table 2.13 Descriptive Statistics for NCDS cohort Model 168](#_Toc174390398)

[Table 2.14 Descriptive Statistics for NCDS model by Dependent Variable 171](#_Toc174390399)

[Table 2.15 Descriptive Statistics Crosstab of NS-SEC for NCDS model 174](#_Toc174390400)

[Table 2.16 Descriptive Statistics Crosstab of RGSC for NCDS model 176](#_Toc174390401)

[Table 2.17 Descriptive Statistics of CAMSIS for NCDS model 178](#_Toc174390402)

[Table 2.18 Model Building Statistics for NCDS model 180](#_Toc174390403)

[Table 2.19 Sequential Model Building Statistics for NCDS model 180](#_Toc174390404)

[Table 2.20 Analytical Model for NCDS 185](#_Toc174390405)

[Table 2.21 Model building statistics of RGSC for NCDS model 194](#_Toc174390406)

[Table 2.22 Model building statistics of CAMSIS for NCDS model 195](#_Toc174390407)

[Table 2.23 Sequential Model building statistics of RGSC for NCDS model 195](#_Toc174390408)

[Table 2.24 Sequential Model building statistics of CAMSIS for NCDS model 196](#_Toc174390409)

[Table 2.25 Sensitivity analysis of social stratification measures for NCDS model 201](#_Toc174390410)

[Table 2.26 Model building statistics of NS-SEC SOC 90 for NCDS model 212](#_Toc174390411)

[Table 2.27 Sequential Model building statistics of NS-SEC SOC 90 for NCDS model 213](#_Toc174390412)

[Table 2.28 Comparison of SOC measures for NS-SEC for NCDS model 216](#_Toc174390413)

[Table 2.29 Model Statistics of RGSC SOC 90 for NCDS model 221](#_Toc174390414)

[Table 2.30 Sequential Model Statistics of RGSC SOC 90 for NCDS model 221](#_Toc174390415)

[Table 2.31 Comparison of RGSC SOC for NCDS Model 225](#_Toc174390416)

[Table 2.32 Model Statistics of CAMSIS SOC 90 for NCDS model 230](#_Toc174390417)

[Table 2.33 Sequential Model Statistics of CAMSIS SOC 90 for NCDS model 230](#_Toc174390418)

[Table 2.34 Comparison of CAMSIS SOC Codes for NCDS Model 234](#_Toc174390419)

[Table 2.35 Simulation Regression Models Using a MAR Principle 259](#_Toc174390420)

[Table 2.36 Missing data patterns for NCDS 262](#_Toc174390421)

[Table 2.37 MI versus CRA for NCDS model 271](#_Toc174390422)

[Table 2.38 Descriptive Statistics for BCS Model 279](#_Toc174390423)

[Table 2.39 Descriptive Statistics for BCS Model by dependent variable 282](#_Toc174390424)

[Table 2.40 Descriptive Statistics comparing NS-SEC by SOC2000 and SOC90 codes for BCS model 285](#_Toc174390425)

[Table 2.41 Descriptive Statistics comparing RGSC by SOC2000 and SOC90 codes for BCS model 287](#_Toc174390426)

[Table 2.42 Model Building Statistics for BCS model 289](#_Toc174390427)

[Table 2.43 Sequential Model Building Statistics of NS-SEC for BCS model 289](#_Toc174390428)

[Table 2.44 Analytical Model for BCS 293](#_Toc174390429)

[Table 2.45 Model building statistics of RGSC for BCS model 302](#_Toc174390430)

[Table 2.46 Sequential Model Building Statistics of RGSC for BCS model 302](#_Toc174390431)

[Table 2.47 Model building statistics of CAMSIS for BCS model 303](#_Toc174390432)

[Table 2.48 Sequential Model Building Statistics of CAMSIS for BCS model 303](#_Toc174390433)

[Table 2.49 Sensitivity analysis of social stratification measures for BCS model 308](#_Toc174390434)

[Table 2.50 Model building statistics of NS-SEC SOC 90 for BCS model 316](#_Toc174390435)

[Table 2.51 Sequential Model Statistics of NS-SEC SOC 90 for BCS model 316](#_Toc174390436)

[Table 2.52 Comparison of SOC measures for NS-SEC for BCS model 320](#_Toc174390437)

[Table 2.53 Model building statistics of RGSC SOC 90 for BCS model 326](#_Toc174390438)

[Table 2.54 Sequential Model Statistics of RGSC SOC 90 for BCS model 326](#_Toc174390439)

[Table 2.55 Comparison of RGSC SOC for BCS Model 330](#_Toc174390440)

[Table 2.56 Model building statistics of CAMSIS SOC 90 for BCS model 334](#_Toc174390441)

[Table 2.57 Sequential Model Statistics of CAMSIS SOC 90 for BCS model 334](#_Toc174390442)

[Table 2.58 Comparison of CAMSIS SOC for BCS Model 337](#_Toc174390443)

[Table 2.59 Missing data patterns for BCS 347](#_Toc174390444)

[Table 2.60 Comparison of CRA, Dummy variable adjustment, and MI models for BCS model 351](#_Toc174390445)

[Table 2.61 Descriptive Statistics of First Destinations 361](#_Toc174390446)

[Table 2.62 Descriptive Statistics of First Destinations by Cohort 362](#_Toc174390447)

[Table 2.63 Modelling First Major Transition with Combined Cohorts (Imputed Models) 363](#_Toc174390448)

[Table 3.1 Descriptive Statistics for Youth's First Destinations 380](#_Toc174390449)

[Table 3.2 Descriptive Statistics for Youth's First Destinations by Cohort 381](#_Toc174390450)

[Table 3.3 Multinominal Logistic Regression model of conditionally imputed pooled dataset investigating youths first destination 383](#_Toc174390451)

###### of Appendix Tables

[Figure 5.1 Coefficient Plot of RGSC model 393](#_Toc172884641)

[Figure 5.2 log odds versus quasi-variance statistics of RGSC for NCDS model 394](#_Toc172884642)

[Figure 5.3 Coefficient Plot of CAMSIS model 395](#_Toc172884643)

[Figure 5.4 Predictive and AMEs of RGSC for NCDS model 395](#_Toc172884644)

[Figure 5.5 Predictive and AMEs of CAMSIS for NCDS model 396](#_Toc172884645)

[Figure 5.6 log odds versus quasi-variance statistics for NS-SEC SOC 90 397](#_Toc172884646)

[Figure 5.7 log odds versus quasi-variance statistics of RGSC SOC 90 for NCDS Model 398](#_Toc172884647)

[Figure 5.8 Trace plot summaries for Economic Activity 399](#_Toc172884648)

[Figure 5.9 Trace plot summaries for Educational Attainment 400](#_Toc172884649)

[Figure 5.10 Trace plot summaries for NS-SEC 400](#_Toc172884650)

[Figure 5.11 Trace plot summaries for Housing Tenure 401](#_Toc172884651)

[Figure 5.12 Log odds versus Quasi-Variance Statistics for BCS model (RGSC) 401](#_Toc172884652)

[Figure 5.13 Predictive and AMEs of RGSC for BCS model 402](#_Toc172884653)

[Figure 5.14 Predictive and AMEs of CAMSIS for BCS model 402](#_Toc172884654)

[Figure 5.15 log odds versus quasi-variance statistics for NS-SEC SOC 90 (BCS model) 403](#_Toc172884655)

[Figure 5.16 log odds versus quasi-variance statistics for RGSC SOC 90 (BCS model) 404](#_Toc172884656)

###### of Figures

[Figure 1.1 Typology of Theories of Structure/Agency 24](#_Toc174390530)

[Figure 0.2 Timeline of Cohorts 43](#_Toc174390531)

[Figure 2.1 The Story of First Transitions by Cohort 72](#_Toc174390532)

[Figure 2.2 Log Odds Coefficient Plot of Pooled Logistic Regression Analysis of Youth's First Transition 152](#_Toc174390533)

[Figure 2.3 Predictive Margins of Educational Attainment by Cohort 154](#_Toc174390534)

[Figure 2.4 Predictive Margins of Sex by Cohort 156](#_Toc174390535)

[Figure 2.5 Predictive Margins of Housing Tenure by Cohort 157](#_Toc174390536)

[Figure 2.6 AMEs and Predictive Margins by Cohort 160](#_Toc174390537)

[Figure 2.7 Coefficient plot of analytical model 189](#_Toc174390538)

[Figure 2.8 Log odds versus Quasi-Variance Statistics for NCDS model (NS-SEC) 190](#_Toc174390539)

[Figure 2.9 Predictive and AMEs of NS-SEC for NCDS Model 190](#_Toc174390540)

[Figure 2.10 Predictive Margins of Educational Attainment for NCDS model 191](#_Toc174390541)

[Figure 2.11 Predictive Margins of Sex for NCDS model 192](#_Toc174390542)

[Figure 2.12 Predictive Margins of Housing Tenure for NCDS model 193](#_Toc174390543)

[Figure 2.13 Coefficient plot comparing all three social stratification models 204](#_Toc174390544)

[Figure 2.14 Comparison of log odds versus quasi-variance statistics of NS-SEC and RGSC measures for NCDS model 205](#_Toc174390545)

[Figure 2.15 Comparison of Predictive and AMEs for each social stratification measure for NCDS model 207](#_Toc174390546)

[Figure 2.16 Coefficient plot for NS-SEC SOC 90 model 218](#_Toc174390547)

[Figure 2.17 Comparison of log odds and quasi-variance statistics for NS-SEC SOC codes for NCDS model 219](#_Toc174390548)

[Figure 2.18 Comparison of Predictive and AMEs of NS-SEC SOC Codes for NCDS model 220](#_Toc174390549)

[Figure 2.19 Coefficient plot for RGSC SOC90 model 227](#_Toc174390550)

[Figure 2.20 Comparison of log odds versus quasi-variance statistics for RGSC SOC Codes 228](#_Toc174390551)

[Figure 2.21 Comparison of Predictive and AMEs for RGSC SOC Codes for NCDS Model 229](#_Toc174390552)

[Figure 2.22 Coefficient plot for CAMSIS SOC90 model 235](#_Toc174390553)

[Figure 2.23 Comparison of Predictive and AMEs for CAMSIS SOC Codes 236](#_Toc174390554)

[Figure 2.24 Coefficient plot comparing all SOC 90 models 237](#_Toc174390555)

[Figure 2.25 Comparative Coefficient plots by SOC constructions of social stratification measures 238](#_Toc174390556)

[Figure 2.26 Comparison of Predictive and AMEs for all social stratification measures for NCDS model for SOC 90 codes 240](#_Toc174390557)

[Figure 2.27 Comparison of Predictive and AMEs of all SOC codes for NCDS Model 241](#_Toc174390558)

[Figure 2.28 Coefficient plot comparing CRA and MI models 269](#_Toc174390559)

[Figure 2.29 Margins plot comparing CRA and MI models 270](#_Toc174390560)

[Figure 2.30 Coefficient plot of main analytical model 296](#_Toc174390561)

[Figure 2.31 Log odds versus Quasi-Variance Statistics for BCS model (NS-SEC) 297](#_Toc174390562)

[Figure 2.32 Predictive and AMEs of NS-SEC for BCS Model 298](#_Toc174390563)

[Figure 2.33 Predictive Margins of Educational Attainment for BCS model 299](#_Toc174390564)

[Figure 2.34 Predictive Margins of Sex for BCS model 300](#_Toc174390565)

[Figure 2.35 Predictive Margins of Housing Tenure for BCS model 301](#_Toc174390566)

[Figure 2.36 Combined coefficient plot of SOC 2000 social stratification measures 310](#_Toc174390567)

[Figure 2.37 Comparison of log odds versus quasi-variance statistics of NS-SEC and RGSC measures for BCS model 312](#_Toc174390568)

[Figure 2.38 Comparison of Predictive and AMEs for each social stratification measure for BCS model 314](#_Toc174390569)

[Figure 2.39 Coefficient Plot of NS-SEC SOC 90 model 322](#_Toc174390570)

[Figure 2.40 Comparison of log odds versus quasi-variance statistics for NS-SEC SOC Codes (BCS model) 323](#_Toc174390571)

[Figure 2.41 Comparison of Predictive and AMEs for NS-SEC SOC Codes for BCS Model 325](#_Toc174390572)

[Figure 2.42 Coefficient plot of RGSC SOC 90 model 331](#_Toc174390573)

[Figure 2.43 Comparison of log odds versus quasi-variance statistics for RGSC SOC Codes (BCS model) 332](#_Toc174390574)

[Figure 2.44 Comparison of Predictive and AMEs for RGSC SOC Codes for BCS Model 333](#_Toc174390575)

[Figure 2.45 Coefficient of CAMSIS SOC 90 model 338](#_Toc174390576)

[Figure 2.46 Comparison of Predictive and AMEs for CAMSIS SOC Codes for BCS Model 339](#_Toc174390577)

[Figure 2.47 Comparison of Predictive and AMEs for all social stratification measures for BCS model for SOC 90 codes 341](#_Toc174390578)

[Figure 2.48 Coefficient plot comparing social stratification measures by SOC 343](#_Toc174390579)

[Figure 2.49 Comparison of Predictive and AMEs for all social stratification measures for BCS model for SOC 90 code 345](#_Toc174390580)

[Figure 2.50 Comparison coefficient plot of CRA and MI models 355](#_Toc174390581)

[Figure 2.51 Comparison of CRA and MI NS-SEC margins 356](#_Toc174390582)

[Figure 2.52 Comparison of Pooled Average Marginal Effects and Predicted Probabilities of NS-SEC by Cohort (CRA on left, MI on right) 366](#_Toc174390583)

[Figure 2.53 Comparison of Pooled Average Marginal Effects and Predicted Probabilities of Educational Attainment by Cohort (CRA on left, MI on right) 368](#_Toc174390584)

[Figure 2.54 Comparison of Pooled Average Marginal Effects and Predicted Probabilities of Sex by Cohort (CRA on left, MI on right) 369](#_Toc174390585)

[Figure 2.55 Comparison of Pooled Average Marginal Effects and Predicted Probabilities of Housing Tenure by Cohort (CRA on left, MI on right) 370](#_Toc174390586)

[Figure 3.1 Predictive and average marginal effects of NS-SEC on youth's first destination by cohorts 387](#_Toc174390587)

[Figure 3.2 Predictive and average marginal effects of educational attainment on youth's first destination by cohorts 388](#_Toc174390588)

[Figure 3.3 Predictive and average marginal effects of sex on youth's first destination by cohorts 389](#_Toc174390589)

[Figure 3.4 Predictive and average marginal effects of housing tenure on youth's first destination by cohorts 390](#_Toc174390590)

[Figure 3.5 Coefficient plots of each category of youth's first destination 391](#_Toc174390591)

[Figure 5.1 Coefficient Plot of RGSC model 393](#_Toc174390592)

[Figure 5.2 log odds versus quasi-variance statistics of RGSC for NCDS model 394](#_Toc174390593)

[Figure 5.3 Coefficient Plot of CAMSIS model 395](#_Toc174390594)

[Figure 5.4 Predictive and AMEs of RGSC for NCDS model 395](#_Toc174390595)

[Figure 5.5 Predictive and AMEs of CAMSIS for NCDS model 396](#_Toc174390596)

[Figure 5.6 log odds versus quasi-variance statistics for NS-SEC SOC 90 397](#_Toc174390597)

[Figure 5.7 log odds versus quasi-variance statistics of RGSC SOC 90 for NCDS Model 398](#_Toc174390598)

[Figure 5.8 Trace plot summaries for Economic Activity 399](#_Toc174390599)

[Figure 5.9 Trace plot summaries for Educational Attainment 400](#_Toc174390600)

[Figure 5.10 Trace plot summaries for NS-SEC 400](#_Toc174390601)

[Figure 5.11 Trace plot summaries for Housing Tenure 401](#_Toc174390602)

[Figure 5.12 Log odds versus Quasi-Variance Statistics for BCS model (RGSC) 401](#_Toc174390603)

[Figure 5.13 Predictive and AMEs of RGSC for BCS model 402](#_Toc174390604)

[Figure 5.14 Predictive and AMEs of CAMSIS for BCS model 402](#_Toc174390605)

[Figure 5.15 log odds versus quasi-variance statistics for NS-SEC SOC 90 (BCS model) 403](#_Toc174390606)

[Figure 5.16 log odds versus quasi-variance statistics for RGSC SOC 90 (BCS model) 404](#_Toc174390607)

###### of Abbreviations

A’level Advanced Level

BCS British Cohort Survey

CAMSIS Cambridge Social Interaction and Stratification Scale

CRA Complete Records Analysis

CSE Certificate of Secondary Education

CTC Central Training Council

ERSC Economic Social Research Council

GCSE General Certificate of Secondary Education

GNVQs General National Vocational Qualifications

ISCO International Standard Classification of Occupations

ITBs Industrial Training Boards

KHB Karlson-Holm-Breen

MCS Millennium Cohort Study

MI Multiple Imputation

MSC Manpower Services Commission

NATFHE National Association of Teachers in Further and Higher Education

NC National Curriculum

NCDS National Childhood Development Study

NDYP New Deal for Young People

NEET Not in Education, Employment, or Training

NS-SEC National Statistics Socio-economic classification

NTI New Training Initiative

NTTF National Training Task Force

NUT National Union of Teachers

NVQ National Vocational Qualification

NWS New Workers Scheme

O’Grade Ordinary Grade

O’Level Ordinary Level

ONS Office for National Statistics

RGSC Registrar General Social Class

RGSC Registrar General’s Social Classes

RSSL Recruitment Subsidy for School Leavers

SOC Standard Occupational Classification

TECs Training and Enterprise Councils

TES Temporary Employment Subsidy

TOP Training Opportunities Programme

TOPs Training Opportunities Scheme

TUC Trade Union Congress

TVEI Technical and Vocational Education Initiative

UKHLS United Kingdom Household Longitudinal Study

WEP Work Experience Programme

WP Work Programme

YES Youth Employment Subsidy

YOP Youth Opportunities Programme

YT Youth Training

YTS Youth Training Programme

YWS Youth Worker Scheme

# Introduction

This thesis employs quantitative analyses to understand how structural inequalities influence young people’s transitions from school-to-work. The analysis uses several longitudinal surveys, including the National Childhood Development Study, the British Cohort Study, the British Household Panel Survey, and the United Kingdom Household Longitudinal Study, to examine the differences within and between cohorts over time. The first survey, the National Childhood Development Study follows individuals born in the UK in one week in March in 1958. The last survey, the United Kingdom Household Longitudinal Survey, began in 2009 following the subsuming of the British Household Panel Survey sample that began in 1991. This thesis covers the latter half of the 20th century to understand the potential changing role of structural inequality and its influences upon youth transition. This thesis explores how the impact of structural inequalities has evolved on youth transitions over time. Whilst the socio-historical context of individuals within each dataset is unique, all individuals across this thesis are encapsulated within the ‘post-Raising of the school leaving age’ – or post-ROSLA – era. One of the constants across all individuals in this thesis is that they all had to continue schooling until at least the age of 16. This constant, grounds this analysis with the knowledge that any differences identified within subsequent analysis of these cohorts cannot be explained away by legal changes to the school leaving age or mandatory schooling.

Extensive research has established that structural inequalities, hitherto defined as including education, sex, housing tenure, and social class, significantly shape youth transitions from school-to-work (Furlong and Cartmel, 1997; Bynner, 1998, 1999; Croxford *et al.*, 2006; Duckworth and Schoon, 2012; Dorsett and Lucchino, 2013; Duta and Iannelli, 2018; Duta, Wielgoses and Iannelli, 2020). Empirical research has consistently demonstrated that structural inequalities contribute to an individual’s school-to-work transition (Jones, 1986; Howieson and Iannelli, 2008; Furlong, 2010; Iannelli and Smyth, 2017).

This thesis contributes to an approach that makes use of gaps in the provision of cohort study data within the sociology of youth through empirical enquiry of school-to-work transitions. First, it provides new empirical evidence analysing the school-to-work transitions of youth, focusing on the nature of structural inequalities influence within and between cohorts. This is accomplished by following a research tradition of creating synthetic cohorts for analyses (Gayle, Lambert and Murray, 2009; Murray, 2011). Through the creation of synthetic cohorts, the youth transitions of individuals in the UK will be studied over the course of half a century, spanning the post-war period up to the start of the millennium.

Second, it builds upon the study of social stratification by deploying sensitivity analyses of social stratification measures to understand if there are any substantive differences in using one social stratification measure over another. Thirdly, it seeks to improve upon classical sociological research in youth transitions methodologically by handling missing data, which is accomplished by discussing and implementing different handling of missing data techniques. All three contributions seek to understand structural inequalities' influence on youth’s first large decision in life, the school-to-work transition following mandatory education. How youth transitions may have altered over time as a consequence of changes within structural inequality is an ancillary question to understand.

In the following chapter, the main themes of the thesis are presented alongside appropriate empirical and social theory literature. The research questions, data, and methods are presented, and the thesis's overall structure is outlined.

## School-to-work transitions in context

Youth transitions encompass what Mayer and Schoepflin (2022) call, key life domains; such as education and preparation for work, as well as active employment, both of which are central to the study of youth transitions. These life domains in the form of school-to-work transitions have a rich sociological tradition within youth research (Clarke, 1978; Raffe, 1984; Bynner, 1998, 1999; Gayle, 1998; Vickerstaff, 2003; Croxford *et al.*, 2006; Brooks, 2009; Iannelli and Smyth, 2017). This research recognises that the socio-historical context under which young people began transitioning into adulthood, has overtime shifted dramatically. These shifts come in the form of: education, the labour market, and economic policies. The youth transitions of 50 years ago are dramatically different from contemporary periods (Murray and Gayle, 2012).

After the Second World War, extended education became more common as a consequence of the raising of the school leaving age to 15 in 1947 – and even more common when the school leaving age was raised once more to 16 in 1972. From the mid-point of the 20th century the youth labour market began to enter a steady decline, with it collapsing in the 1980s. Economic policies also began to shift, as a result of exogenous factors affecting the educational and economic landscape. As a result, traditional apprenticeships were phased out due to their heavy reliance of heavy industrial manufacturing in favour of service based short-term youth training schemes. All these changes were part of a much broader economic and socio-political shift towards post-industrial thinking that dominants the current consensus concerning British society (ibid). These developments all collectively contribute to the ‘changing times consensus’, a term coined by Gayle, Lambert, and Murray (2009). The changing times consensus refers directly to the pervasive changes in school-to-work transitions during this period. Change is the defining factor of school-to-work transitions in the latter half of the 20th century. This change is an amalgamation of shifts in educational policy and administration, change as a result of exogenous economic shocks disrupting established labour market practices, and change in state economic policies leading to a re-structuring of youth labour market entries and reformation of apprenticeship based programs like the New Deal for Young People (NDYP) (Olle, 2022).

The transition from school-to-work and into adulthood lacks a single, universally agreed upon definition. Whilst many influenced by life course theory use terms such as ‘youth phase’ (Elder, Johnson and Crosnoe, 2003; Bynner, 2005). Others call this stage ‘emerging adulthood’ (Arnett, 2000, 2006). Whilst others still call simply call this stage ‘young adults’ (Furlong and Cartmel, 2007). There is no clear-cut definition of where childhood, youth, and adulthood start and end. Life course theory views youth transitions as age-graded boundaries, where each stage, including education and work, represents a critical life domain that evolves over time (Elder, 1994; Mayer, 2009). Life course theory provides the most robust and holistic definition of youth transitions and will be employed going forward.

Previous research on British school-to-work transitions has focused either on a descriptive analysis of spells of transition (Schoon, Ross and Martin, 2009; Schoon, 2012; Anders and Dorsett, 2017) or on a comparison of change between cohorts from different time points (Bynner, 1998, 1999; Bynner and Ferri, 2003). The latter of this research has used the British birth cohorts – the National Childhood Development Study in 1958 and the British Cohort Study in 1970. These Birth cohorts allow for easy comparisons of school-to-work transitions at different socio-historical time points due to much of the survey data being harmonised for comparative use. Unfortunately, no birth cohorts exist from 1970 to the end of the 20th century. The Millennium Cohort Study (MCS) began in 2000 and is a promising birth cohort study that may continue the traditions set out by prior cohort studies of its nature, however respondents are too young to provide youth transitions data. Even if data were available for the MCS, that still does not eliminate the data gap from 1970-2000. This gap in birth cohort data from 1970 onwards presents a significant challenge to studying the evolution of school-to-work transitions in a contemporary context. This thesis addresses this gap by using contemporary statistical techniques to create synthetic cohorts using longitudinal household panel data for analysis.

## Social Theory

The study of school-to-work transitions consistently highlights two interrelated themes: 1) the influence of structural factors, and 2) the role of individual agency. The choices and decisions made by young people are shaped by structural forces that are in turn influenced by the social context of which they exist. Both structure and agency have a long tradition within the sociology of youth and sociology more broadly. Different sociological theories have offered various, often conflicting, interpretations of how structure and agency interact, ranging from deterministic models that emphasis structural constraints, to those that highlight individual autonomy. This section attempts to return to that theoretical stage, firstly to address the relevant theories related to the sociology of youth specifically, but secondly, to make it clear which theoretical orientation will take centre stage within this thesis. The sociology of youth has many theories that focus on the role of structure and agency: are transitions on a train track, railroading individuals to a pre-determined destination? Or are individuals free to choose as they wish when they wish? While these examples represent opposing extreme ends of the spectrum, they underscore the importance of revisiting the social theories that form the foundation of youth sociology. In the sociology of youth, it is essential to consider how changing social processes, shaped by varying socio-historical contexts, influence the transitions young people experience. The nature of socio-historical context in of itself may have consequences for youth transitions. The structure/agency problem is central to a study of youth transitions and will be further developed in this section.

The structure/agency problem is one of sociology’s most pressing matters of social theory. The relationship between structures and their constraints and enabling influences on agentic action has been a key focus for a range of social theories that focus on ideas of choice and opportunity. Evans (2007), whilst developing their concept of ‘bounded agency’ also developed a typology to understand the divergences in social theory that purports to analyse the inter-relationships between structure and agency. Evans’ typology is particularly valuable in categorising the diverse social theories relevant to understanding the balance between structure and agency within youth transitions research. Not all social theories are relevant to the study of youth transitions – only theories that offer nuanced insights into how structural forces and individual agency intersect are particularly relevant and considered here. I have updated Evans’ typology for the sake of this thesis in Figure 1.1.

A diagram of a triangle

Description automatically generated

Figure ‎1.1 Typology of Theories of Structure/Agency

This typology has three dimensions. The first relates to the primacy structure versus agency divide. Certain theories give primacy to structure over agency, such as Bourdieu’s theory of habitus (Bourdieu, 1989, 1993, 2013), whereas others argue that structures in themselves have eroded in favour of agentic primacy (Baudrillard, 1988). Gidden’s structuration theory (Giddens, 1989) is positioned within the middle of this dimension as his theory expresses an interdependence of structure and agency, whereas Beck’s Individualisation thesis (Beck, 2002, 2014) purports individuals engaged in the construction of their own biographies is placed closer to agentic primacy – but not as far on the scale as structural erosion theory. The second dimension emphasises types of control processes. These control processes are dichotomised into internal control – whereby internal processes of the acting individual relate to the external environment, or external control – whereby external limits are placed upon internal processes. On the external end of this scale are theories of the life course represented as age-graded social trajectories subject to changing external conditions (Elder, 1994, 1995:48; Elder, Johnson and Crosnoe, 2003; Hitlin and Elder, 2007). The third and final dimension focuses on the nature of social relationships and how they can be converted or reproduced for individual and collective actions (Evans, 2007: 16). The emphasis on social reproduction is the main focus of original work on rational action theory (Goldthorpe, 1998).

This typology is helpful in differentiating different social theories of structure and agency. This thesis centres on a handful of the social theories expressed here that are integral to an effective understanding of the sociology of youth and understanding the statistical relationships that develop over the course of a quantitative analysis of youth transitions across cohorts. A primary focus will be on Gidden’s Structuration. (1989) Beck’s Individualisation (2002), Elder’s Life Course (1994), and the mid-range theories of Bounded Agency (Evans, 2007) and Structured Individualism (Roberts, Clark and Wallace, 1994).

This section on social theory will start with an explication and justification of the use of the theory of the life course as developed and deployed by the likes of Elder (Elder, 1994, 1995; Elder, Johnson and Crosnoe, 2003; Hitlin and Elder, 2007) and Mayer (Mayer, 2004, 2009; Brückner and Mayer, 2005; Diewald and Mayer, 2008; Mayer and Schoepflin, 2022) – highlighting the particular importance of socio-historical context to construct life biographies and its importance for cross-cohort comparisons. Then the theory of Structuration will be discussed, with the original theory explained (Giddens, 1989), critiqued (Archer, 1995; Mouzelis, 1997; Healy, 1998) and reformulated as a justified theory to understand social stratifications influence upon individual action in terms of youth transitions. The theory of Individualisation (Beck, 1992, 2002, 2014; Beck, Giddens and Lash, 1994) will be introduced as a critique of structuralism but ultimately presented as a flawed alternative alongside the theories of New Structuralism (Saunders, 2003, 2021; Devine, 2017). Finally, the reformulated Structuralist theory will be placed alongside the mid-range theories of Structured Individualism (Roberts, Clark and Wallace, 1994) and Bounded Agency (Evans, 2007), with the latter two being presented as justifiable theoretical orientations to understand the stratifying influence of structural inequalities upon individual agency and how that directly relates the empirical study of school-to-work transitions. This section is an attempt to justify the use of certain social theoretical orientations such as the life course, structuration, and bounded agency to synthesise an adequate explanation of potential statistical phenomena. The social theory presented here enables a comprehension of the changing context of UK society and its potential impacts upon youth transitions.

The Life Course

The life course approach provides a dynamic framework for understanding youth transitions by viewing them as part of an individual’s age-graded social biography (Elder, 1994). Unlike static models that capture only a moment in time, the life course approach considers individuals within a web of continuously evolving temporal contexts that shape their choices and opportunities. Life course theory naturally lends itself to longitudinal research over the more static snapshot cross-sectional studies. It incorporates the changing processes and influences that impact an individual’s choices and opportunities during transitions during the youth stage. Given the longitudinal nature of the data employed in this thesis, the life course approach is particularly well-suited for analysing how structural inequalities impact youth transitions over time.

The life course approach has established itself as a substantively significant research paradigm within the last few decades. (Elder, 1994). The life course refers to the multilevel phenomena where individuals navigate social trajectories through structured institutional pathways, that shape their developmental processes (Elder, 1994). These structured pathways provide for a framework to analyse how structural inequalities – such as sex, housing tenure, and social class – shape the opportunities and outcomes in school-to-work transitions (ibid). These trajectories took the form of work, family, and housing transitions. Such transitions are always historically and temporally located, giving them specific form, and meaning (ibid). In addition, each individual life history and trajectory is bound through an interdependence of life domains. (Mayer, 2009). This means that outcomes within one domain (e.g., school) are interrelated with the outcomes and behaviours of other domains (e.g., work). The structured pathways within the life course support an analysis focusing on inequalities concerning race, class, gender, and other structural aspects of social life. (Bernardi, Huinink and Settersten, 2019). The life course approach is uniquely poised for a detailed study of youth transitions. Youth transitions, by their very nature, detail pathways of trajectories that individuals enter at specific points in their lives that are ultimately influenced and dependent upon structural inequalities.

This thesis will study the impact of structural inequalities on an individual’s ‘choice’ and opportunity post-mandatory schooling. The term ‘choice’ is problematised to highlight how it is often contrasted by structural and contextual factors, making it a contentious concept in the context of youth transitions. The choice is dependent upon many aspects of an individual’s situation within the life course. Indeed, for a child born to wealthy parents, the choice of whether to send said child to private school or not is a very real one. For a child born to parents on the poverty line, the ‘choice’ has already been made. The parents have little economic capital to leverage, and so there is a strong likelihood that the child will not go to private school in this instance. In this respect, ‘choice’ is not the same as a ‘choice’ being made by a family with an advantaged social position. Whilst there are avenues for a working-class child to attend private schools – such as a bursary or grant, or a wealthy distant relative, etc., these opportunities are few and far between, and the most likely result for a child of working-class origins is to attend a comprehensive school. Choice, when it is used in the context of youth transitions and indeed this thesis, must come with it an asterisk, understanding that whilst there is almost always a possibility of an individual being able to do something, the probability of that choice occurring is in fact influenced by their given circumstances. This thesis adopts a definition of choice that is derived from the ‘principle of agency’: ‘’individuals construct their own life course through the choices and actions they take within the opportunities and constraints of history and social circumstances’’ (Elder, Johnson and Crosnoe, 2003). This concept aligns itself with the idea of ‘bounded agency’, where individual choices are shaped and limited by the structural contexts in which they are made (Evans, 2007). Bounded agency is a concept that argues that the agency of the individual is situational and bound to the circumstances of place and time (Bernardi, Huinink and Settersten, 2019).

Understanding how structural inequalities influence post-mandatory schooling decisions is crucial for assessing their long-term impact on life chances across the life course. Functionally, the study of youth transitions is the study of the life course; education systems, occupation, and labour markets that constitute life domains onto which the life course manifests (Mayer, 2009). Mayer argues that these life domains are organised around three major life stages: the phase of education and preparation to work, the phase of active employment, and the phase of postretirement years (Mayer and Schoepflin, 2022). The study of school-to-work transitions constitutes two of these life domains. These life stages vary in timing and sequencing depending on socio-historical context (Shanahan, 2000), which must be acknowledged for adequate sociological inquiry.

A life course perspective provides the necessary theoretical orientation to investigate youth transitions within their own cohorts. Beyond this, the life course also facilitates comparison across cohorts to study how different cohorts have responded to the consequences of their prior life domains.

Whilst the life course forms the foundation for subsequent analysis, the purpose of the analysis itself is the investigation of structural inequalities in school-to-work transitions. Structural inequalities are viewed through the lens of the role of structure. Although post-modernist scholars such as Lyotard (1984) and Baudrillard (1988) argue that structural analysis is outdated, this thesis maintains that structural inequalities remain central to understanding life chances, as evidenced by persistent social inequalities based upon sex, housing tenure, and social class (Furlong and Cartmel, 2006).

Structuration

Giddens’ theory of structuration provides a framework for understanding how structures such as: sex, housing tenure, and social class shape and are shaped by, individual action (Giddens, 1989). Structuration argues that structural factors like still play an essential role in shaping the lives of individuals and are indeed determinants for the individual pursuing the ‘imperative of living a life of one’s own.’ (Beck, 2002), though also emphasis the role of individual agency in navigating and transforming these structures. The life course as a social theory enables us to understand individual action through age-graded social biographies (Elder, 1994), but the social theory of Structuration enables us to also understand the construction of structures themselves and how that influences individuals agency and vice versa.

Structuration theory is built upon the premise of the duality of structure (Giddens, 1989). The duality of structure suggests that structures are both the medium and outcome of social practices, this means that while structures shape individual actions, those actions simultaneously reproduce or alter the structures that inform action. Though it is better named the duality of structure and agency, given the concept's refusal to give either primacy over the other. Unlike traditional structural theories such as Parsonian Structuralism, which views structures as external and beyond individual control, structuration theory argues that structure and agency are mutually constitutive. Contrasting the theoretical concepts of the duality of structure with that of structural dualism offers a more sophisticated comparison of structures that moves beyond the dominance-of-one approach that traditional theories produce. The former of which is a foundational pillar of structuration theory stating that the "structural properties of social systems exist only insofar as the forms of social conduct are reproduced chronically across time and space" (Giddens, 1989: xxi).  Giddens introduces the concept of practical consciousness to explain how individuals, through routine actions, perpetuate or change social structures. Practical consciousness refers to the tacit knowledge individuals use in their everyday activities, guiding their actions beyond what can be explicitly articulated. Said practical consciousness extends the reflexivity of the agent beyond the mere discursive ability to state why and what they do. Practical consciousness becomes a site of knowledgeability of an agent's ability to know why and what they do (Giddens, 1989).

Giddens emphasises that social systems are inherently tied to time-space relations. An altering of social relations also alters structure, as the duality of structure "is a medium and outcome of reproduction of practices" (Giddens, 1979: 5). The interdependency of agents and structures mutually engaging in and enacting social systems is foundational to the theory of structuration. While the interdependence of structure and agency in structuration theory complicates the identification of clear causal based relationships, it offers a nuanced perspective that captures the complexity of social dynamics and functions.

Critiques of Structuration

Traditional structuration theory does suffer from critiques, primarily concerned with its definition of structure and duality (Healy 1998). The theory itself still provides utility when attempting to explain youth transitions and can be conjoined to other social theories to explicate social processes. Gidden's social reproduction across space-time presupposes the reflexive monitoring of the agent’s participation in social activity. The knowledgeability of the agent is due to this, always bounded by the space-time contexts of existing social relations. This particular point made by Giddens allows structuration theory to ameliorate itself with a theory of Bounded Agency (Evans, 2007). Both theories argue that the agent and their decisions are bounded to a specific socio-historical context.

Individualisation and New Structuralism

A retort to the theory of Structuration (and as such its re-conceptualisations) comes in the form of the theory of individualisation. Individualisation argues that in place of these ‘collective guides’ (Gayle et al., 2009), individualised identities that have greater scope beyond the mere structures (Murray, 2011) they inhabit and can create complex and subjective lifestyles that deviate from the much more rigid structures defined through Structuration (Gayle et al., 2009). This process of Individualisation has been the result of specific historical developments that have resulted in the loss of traditional structural support networks, which has meant that individuals have had to rely upon their own self to guide them through the risks and opportunities of society (Beck, 1992). For example, whilst the extension of mandatory schooling has brought about a standardisation of schooling practices for individuals, the increasing levels of credentialization within the labour market have presented schooling as a highly individualised place of choosing and planning one’s own life course situation (ibid). The individual is required to incorporate the reflexive knowledge of the conditions and prospects of modernity that cut across traditional structural lines and, in this way, become an agent of the reflexive modernisation process (ibid).

The fundamental component of Individualisation is a competing theory of structure/agency relations that argues that structures are becoming detraditionalized and dissolutioned from their particular milieus; this is, so says Beck revealed by the increasing levels of difficulty in interpreting empirical sociology in the form of class and stratification research (Beck, 1992). It is within this difficulty that the individual is seen as the reproduction unit for society in the world of life – an agentic primacy.

If the individualisation thesis were correct, it would demonstrate itself empirically and repeatably. However, as Gayle, Lambert and Murray (2009) found, the thesis’ strong claim against structures is not to be born out within the data. Pathways toward transition may have altered and even, in some cases, become more complex, but that does not mean there is support for ‘detraditionalization’ (ibid) or the death of structures. There has always been an element of navigation and choice within youth transitions. However, in the past, the range of choice may have been narrower, thus owing to a more homogenised pathway for those in past contexts (Goodwin and O’Connor, 2005). Individualisation fails to account for structural elements' still apparently strong influence on a young person. A vital part of this thesis is to dig much deeper into the story of Structuration versus Individualisation. This is accomplished through a cross-cohort comparison of structural inequalities and their influence on individual choice and opportunity. A weakening of structural influence may not definitively prove an Individualisation thesis, but it would certainly add credence to the argument itself.

Whilst Individualism critiques Structuration from the point of agentic primary, the theory of New Structuralism critiques Structuration from the point of structural irrelevance. The theory of New Structuralism states that whilst structures are important in understanding the role of individual action, the traditional structures reported in a theory of Structuration such as social class and sex are inadequate mechanisms of individual action in a post-industrial modernity (Saunders, 2003, 2021; Devine, 2017). In this, New Structuralism and Individualisation both readily agree with the degradation of traditional structural roles and influences. Where New Structuralism departs form Individualism however is that it argues that new social cleavages have developed in the wake of this new modernity. Instead of structures of social class and sex being most important in influencing and determining life chances and opportunities in the form of individual action, it is instead these new consumption cleavages such as housing tenure that operate at the level of influence social class and sex once did in industrial society. Unfortunately, just as with strong versions of Individualism, the New Structuralist thesis doesn’t hold up to empirical critique (Franklin and Page, 1984; Hamnett and Mullings, 1992).

Structured Individualism and Bounded Agency

Proponents of structuration, such as Giddens (1989), appear not to appreciate the increasing complexity levels placed upon individuals, leading to potentially heterogeneous outcomes even if the effect of structural inequalities remains strong. In this case, it is best to call for a ‘Structured Individualism’ thesis (Roberts, Clark and Wallace, 1994). One that recognises in a risk society, that whilst pathways are different and numerous, as empirical data demonstrates (Roberts, 2003), they are still heavily influenced by the structures of society (Gayle, Lambert and Murray, 2009). A life course perspective also advocates for a ‘Structured Individualism’, though calls it an ‘agency within structure’ approach to sociological reasoning (Diewald and Mayer, 2008). Whilst it is important to recognise the structural influences upon individual choice and opportunity, it is crucial to treat the individual as an active agent in shaping their biographies. This deters a deterministic theoretical orientation whilst maintaining that some individuals will structurally have more agentic opportunities based on power relations (Hitlin and Johnson, 2015; Schmitt, 2021). Contemporary social theory has also called for a theory of ‘Bounded Agency’ (Evans, 2007) which follows a very similar theoretical orientation to that of Structured Individualism. A theory of Bounded Agency advocates for a conceptualisation of the agent as an active participant in the process in which past actions and routines are contextualised within their given socio-historical contexts whilst future possibilities are envisaged within the constraints and enabling possibilities of the present (Evans, 2007). This theory of bounded agency improves slightly upon a theory of structured individualism by referencing the space-time dynamics of structural influences upon individual agency more explicitly than the latter theory. It points out that in environments of highly structured activity, there is a reduced possibility of individualisation and in this context the potential consequences of that individuals’ actions are placed at the feet of the structures rather than the individual. Bounded Agency provides an environmental grounding to the arguments of Structured Individualism that allows an explanation of behaviour based on structure/agency, internal/external frames of reference, and internal/external actions (Evans, 2007).

Overall, the literature stresses the relevance of contextual factors that highlight the importance of individual agency (Steiner et al., 2021). Whilst there is no definitive social theory that explains all social phenomena regarding youth transitions, a re-interpreted theory of Structuration that is coupled with theories of the life course and bounded agency provides an adequate theoretical orientation going forward. A requirement of any analysis of youth transitions is to assess the strength of these various social theories and provide appropriate arguments to the empirical support of one or more social theories presented in this section.

## Social Stratification

Social stratification is the economic and social inequality that individuals experience within society. All societies have experienced some level of unequal rewards or resources that create a stratified society. Grusky (1994: 3) argues that the ‘’task of contemporary stratification research is to describe the contours and distribution of inequality and explain its persistence despite modern egalitarian or anti-stratification values”. The degree of inequality or stratification within a society is determined by the dispersion or concentration of assets across a given population (Grusky, 1994). Manifestations such as religion, government, wealth, labour technical knowledge etc are environments in which stratification reifies itself throughout a society (Davis and Moore, 1994). This reification of stratification can come in the form of social class, racial or gendered based inequality. Within contemporary social stratification research, the view that a multidimensional (or intersectional) approach in taking into account all status group memberships is the most appropriate way to understand human behaviour under a stratified system (Grusky, 1994).

### Changes in the social stratification structures

Within capitalism, people are uprooted in successive waves and are loose from the structures of previous time-space contexts. These waves are formed through exogenous economic shocks in the form of recessions, technological and cultural change, and restructuring of the labour market in response to said change. These shocks, through waves of change, uproot the rules and resources that dictate society through structural formations of the family, occupations, neighbourhood, and family. This has led to the increased navigation of risks through individualised identity whilst newly formed structures re-formulate themselves in response to said change.

The resultant shocks and uprootedness experienced by those born in the 1950s and 1960s correspond with a shifting educational reform and focus on elongated periods of schooling (Leuze, 2010). This educational reform, which initially started in 1944, is accompanied by a dependency on education to navigate the newly formed structures of modernity successfully. Increasingly, more and more groups realise that to have a chance at successfully navigating the risks these structures of modernity produce, they have to maintain a prolonged stay within education for the credentials that they provide (Beck, 1992).

These changes to the rules and resources that govern the structures of society do not stop at the realm of educational expansion. Cultural changes have resulted in women entering the workforce en masse. The liberalisation of gender relations with respect to reform within divorce law (allowing no-fault divorces) alongside the rise of women entering the labour market has provided women with economic independence (Smith, 1997). Paid maternity and paternity leave have also provided certain securities for women to protect their employment status post-birth (Canaan *et al.*, 2022). Technological innovation has also provided a ‘contraceptive revolution.’ (Westoff and Ryder, 2015) that allows women to control to a greater extent if and when they wish to have children, as well as the development of household utilities that ‘deskill’ housework (Beck, 1992).

Whilst the demographic liberation, deskilling of housework, contraceptive technology, and participation within education have led to the liberation of women among traditional and feudalistic gendered fates, the fact remains that the equalisation of gender relations cannot be created within institutional structures that presuppose their inequality (Beck, 1992). As the rules and resources of structures alter, opportunities and constraints on choice also alter for the biographical dimension of the agent but remain influenced by the structural artefacts of prior structurally composed epochs, such as sex.

## The British Education System

The start of the latter half of the 20th century has been described as ‘capitalism’s golden age’ (Birnbaum, 2002). This golden age saw education spending increase to vast amounts – between 1951 and 1975 spending on education rose from 6.5 per cent to 12.5 per cent of public expenditure (Jones, 2016). Spending was not the only change in relation to the education system. The Education Act of 1944 (followed by a Scottish equivalent in 1945) attempted to remove the fundamental inequalities inherent in the institution of education in Britain. The reform act did two fundamental things that changed the structure of education and youth transitions. The first was the re-structuring of ‘streams’ of schools – grammar, secondary moderns, and technical. This was established based on the recommendation of Sir William Spens White Paper (*Spens Report*, 1938). The second was the raising of the school leaving age to 15 and then to 16 in 1972. This reform was meant to remove barriers to entry for those children from different social origins, though the tripartite system on which the 1944 reform act was built structurally divided those youth depending upon academic ability.

In 1965, 92 per cent of students in state secondary education were in schools organised along tripartite lines, by 1976, comprehensive schools accounted for 76 per cent of students in state secondary education (Jones, 2016). A comprehensive school was one that had no selective criteria based on academic performance. This stark flip from tripartite to comprehensive schooling was the influence of the 1964 Labour general election victory and more specifically the Secretary of State for Education Anthony Crosland who stated ‘’If it’s the last thing I do, I’m going to destroy every last fucking grammar school in England. And Wales. And Northern Ireland’ (Kogan, 2006).

In 1961, 73 per cent of students in England and Wales left school without ever having attempted a public examination, and over 90 per cent of Scottish school leavers left at age 15 (ibid). The restructuring of the British economy saw an increased demand for certification and credentialization (Jones, 2016). This coupled with the Crowther Reports (*Crowther Report Volume I*, 1959) findings that ‘education is a vital part of the nation’s capital investment’ and the Newsom Report’s (*Newsom Report*, 1963) belief that investment in education is a compelling economic argument also saw the view of education transformed into an economic argument.

Following these increased demands the Certificate of Secondary Education, an examination below the General Certification of Education was introduced in 1963 (with the first examinations being held in 1965)– though it took 10 years for an equivalent examination to be established in Northern Ireland (Jones, 2016).

The changing structure of British society in the 1960s and 1970s had an impact on continuing education for 17-year-olds. For those in England in 1966 the proportion of 17-year-olds still in school was 12 per cent. By 1977 this number had risen to 17 per cent. This again rose to 24 per cent in 1979 and over 60 per cent in 1994 (Jones, 2016).

In 1988, a second Education Reform Act was introduced following the introduction of the General Certification of Secondary Education and the ending of the two-tier examination system. The 1988 act established a national curriculum and ended gendered segregation to certain academic subjects like maths and science (Jones, 2016).

The British education system has witnessed dramatic restructuring since the end of the Second World War. The major changes and developments have been highlighted, but given that England and Wales, Scotland, and Northern Ireland all have different educational institutions, it is simply not possible to detail all changes from the 1950s onwards. The key takeaway from the reform of the British educational system is one of expansion and (an attempt) to tackle structural inequalities within the foundation of the provision of education.

## The British (Youth) Labour Market

This section provides a brief overview of the changing nature of the labour market from 1958 onwards. This section seeks to provide an overall timeline to refer to when analysing between cohort trends. Discussions of the youth labour market are orientated around young people bounded by the Statutory School Leaving Age of 16 and the attainment of legal adult status at age 18 (Deakin, 1996). However, some vocational education and training from the 1960s onwards has targeted youth from as young as age 14, and modern apprenticeship schemes end at age 20. The youth labour market discussed hitherto will concern primarily labour market activities post-mandatory education - though discussion of training schemes prior to this will be referenced where relevant.

Whilst there exists some support for the notion that the youth labour market is pegged to that of the general labour market, and reacts in a similar way to general shifts in the level of demand (Makeham, 1980; Raffe, 1984). Makeham's analysis of the youth labour market suggests a certain degree of gender sensitivity with the general labour market based upon unemployment. Between the years 1959-1977, for every 1 per cent increase in overall male unemployment, male youth unemployment rose 1.7 per cent, for women, every 1 per cent increase saw female youth unemployment rise by 3 per cent (Makeham 1980).

From the birth of the NCDS cohort to the start of the 21st century the British economy and labour market experienced a drastic period of restructuring. A simultaneous decline in heavy manufacturing and the youth labour market in turn saw a rise of a service based economy and soft skill labour (Bynner, 2012). The period covering an analysis of the NCDS, BCS, and UKHLS cohorts covers a period of over 50 years. During this period the British economy suffered seven major recessions and seven instances where the governing party changed between the Conservatives and Labour as seen in figure 1.2.

A close-up of a graph

Description automatically generated

Figure ‎1.2 Timeline of Cohorts

[under construction, will be changing when I get the exact UKHLS cohorts]

The contributing factors of the demise of heavy industry, the collapse of communities, and technological white heat of modes of production led to the gradual demise of youth labour within the UK (Bynner, 2012). This collapse of youth labour produced an element of instability within young people’s lives. Individuals from the late 1970s onwards were faced with a new labour market landscape comparative to their parents. Declining manufacturing and increased automation impacted certain areas of the UK more than others. The 1970s recessions that devastated heavy manufacturing located within the North of England (Hamnett, McDowell and Sarre, 1989) not only affected families in manual occupations, but also affected those families living in specific localities. The measures taken by the Conservative government during the 1980s recession that soon followed directly contributed to the further decline of manufacturing jobs in the UK using monetarist polices that attempted to lower inflation. The result saw Northern Ireland have an unemployment rate of 20 per cent, and Scotland and Northern England had an unemployment rate of over 15 per cent (*BBC*, 1982). Post-recession, half of all jobs created between the years 1983-87 were made in the South-East of England (Hamnett, McDowell and Sarre, 1989).

The large-scale increase in youth unemployment during the 1970s is primarily related to the declining role of traditional manual occupations that were mainly provided through apprenticeship programs for young people (Deakin, 1996) . In other words, the declining general demand for manual occupations had a direct effect on the declining levels of opportunities for youth training, which as a consequence has an impact upon youth unemployment. The routes for young people from the 1970s were constraining. The number of apprenticeships in manufacturing industries was halved between 1965 and 1982 from 243,000 to 123,000. In 1990 this number reduced further to just 54,000 (Deakin 1996).

The youth labour market collapsed, heavy manufacturing declined, and the rise of technology gave rise to increasing automation. On top of this, exogenous shocks disproportionately affected the labour market that youth were most likely to enter, as well as continuing to affect manual labour occupations. The distribution of these exogenous shocks was primarily located in the North of England, Wales, Northern Ireland, and Scotland. Collectively, these factors also resulted in a decline of a key avenue of youth transition: the apprenticeship. It was eventually replaced by the Youth Training Scheme (YTS) (Droy, Goodwin and O’connor, 2019). The YTS scheme sought to provide youth with training schemes in a new service-based labour market. The training schemes were unattractive and provided risky employment prospects (ibid).

## Research Questions

The main aim of this thesis is to develop a detailed understanding of the influence that structural inequalities have upon an individual’s school-to-work transition post-mandatory schooling. The primary mode of exploring this topic is through the development and application of statistical models to large-scale complex social science survey data.

The thesis itself is split into three distinct though interrelated parts. Part 1 is a historical study of youths' first transition using the National Childhood Development Study, British Cohort Study, and the British Household Panel Survey and the United Kingdom Household Longitudinal Survey. A simple logistic regression with a binary outcome of continuing schooling versus not continuing schooling will be provided to establish an initial simple analysis of school-to-work transitions. Within this part detailed descriptions of model selection, sensitivity analysis, and how this thesis manages missing data is provided. The overarching research question for Part 1 is:

1. What are the patterns of social inequality in youth transitions?

This primary research question will follow through into parts 2 and 3 of the thesis. Part 2 of the thesis will move on to a study of the first destinations on youths moving from school to work using the same datasets. This section will provide a more granular detail of school-to-work youth transitions by using a multinominal logistical regression of the possible destinations youth from each cohort enter after mandatory schooling.

Parts 1 and 2 will combine to answer research questions two and three:

1. How have patterns and trends in youth transitions changed over time?
2. How have the social processes that underpin youth transitions changed over time?

Finally, part 3 directly explores a life course-based analysis first by conducting an optimal matching analysis and subsequent cluster analysis to produce a descriptive picture of young people’s economic activity trajectories from the end of mandatory schooling to age 23. After this descriptive picture is realised, the clusters from the cluster analysis will be used in a multilevel model using clusters at level two with individuals at level 1 to understand the impact school-to-work trajectories have upon an individual’s later life course. This section of the analysis will answer the fourth research question:

1. How can youth transitions be more comprehensively understood within a life course perspective?

A fourth part to the thesis will function as an overall comparison and conclusion of findings to reflect on the answers of each of the four research questions.

## Data

The relationship between structural inequalities and individuals’ school-to-work transitions is examined using large-scale, nationally representative data collected from two birth cohort studies and two household panel surveys. Birth cohorts offer a practical solution to the analysis of youth transitions through longitudinal data collection strategies. Birth cohorts offer a systematic sampling strategy that solely focus upon individuals born within a specific time frame enabling an easy isolation of age effects. The NCDS cohort were born in 1958 and provide one of the earliest opportunities to study a post-war cohort. The BCS cohort were born in 1970 and as a dataset is often used in combination with the NCDS to analyse trends over time. There is a lack of birth cohorts for the rest of the century. An alternative is to construct synthetic cohorts from other surveys. The UKHLS provides an opportunity to do just that, the timeframe of the BHPS and UKHLS allows for the construction of synthetic cohorts covering the period of the 1990s – a time that is lacking in youth studies literature.

The first birth cohort study is the National Childhood Development Study (NCDS) (Power and Elliott, 2006). The second birth cohort study is the British Cohort Study (BCS) (Elliott and Shepherd, 2006). The first household panel survey is the British Household Panel Survey (BHPS) (Taylor *et al.*, 2018). The second household panel survey is the United Kingdom Household Panel Survey (UKHLS) (Buck and McFall, 2011). The following sections outline the relevance of each database for inclusion in this thesis, the issues and considerations that result from working with birth cohorts and household panel surveys, and an overview of the proposed statistical modelling and methods used within this thesis.

### National Childhood Development Study

This work will use the National Child Development Study using[[1]](#footnote-1) (University of London, 2023c, 2023b)[[2]](#footnote-2). The NCDS is a nationally representative birth cohort study that uses a systematic sample to follow the lives of individuals from England, Wales, and Scotland from birth (Power and Elliott, 2006). The NCDS has an initial sample of 17,415 participants using a cross-sectional sampling design to collect participants from birth within the week of 3-9 March 1958 (Shepherd, 1995). Originally designed to examine the social factors associated with perinatal mortality, the purpose of the NCDS gradually extended to studying other aspects of individuals' lives as they entered adulthood.

The sources of information and methods of data collection for each sweep of the NCDS altered primarily depending on the age of the cohort. The original birth sweep completed as a Perinatal Mortality Survey questionnaire was completed by a midwife who interviewed the mother and consulted their medical records. The first follow-up sweep consisted of a parental interview, a medical questionnaire completed by a medical officer, an educational questionnaire completed by the head teacher and class teacher at the cohort members' school, and finally, a test booklet completed by the cohort member in the school. The second and third follow-ups were identical in scope and survey instrument to the first follow-up. The third follow up differed slightly in that details of examination performance by members of the cohort were obtained in 1978 by writing to schools which study members were known to attend at the time of the 1974 follow-up (Shepherd, 1995). The fourth follow-up was the first instance of the cohort member being an adult, and thus, they took primary control of answering the survey instruments (ibid). This sweep consisted of a cohort member interview that was undertaken by a market research interviewer and supplemented by the 1971 and 1981 UK censuses. The fourth follow-up also provided a feasibility study in 1978 to assess the ability to track and trace cohort members now they had entered adulthood, the study found after attempting to contact and trace a five per cent random sample of those involved in one or more NCDS sweep that it was possible to find the majority of those involved in the NCDS (Shepherd, 1995).

For a full breakdown of information sources for each wave of the NCDS included in subsequent analysis, refer to table 1.1. The methods for collecting information for the BCS included face-to-face interviews, proxy interviews, telephone interviews, self-complete questionnaires, assessments, and medical measurements.

Table 1.1 Follow-ups and information sources for NCDS

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year | 1958 | 1965 | 1969 | 1974 | 1981 |
| Age | Birth | 7 | 11 | 16 | 23 |
|  | Mother | Parents | Parents | Parents | Subject |
|  | Medical | Medical | Medical | Medical | Census |
|  |  | Tests | Tests | Tests |  |
|  |  | School | School | School |  |
|  |  | Subject | Subject | Subject |  |
|  |  |  |  | Census |  |

The sampling strategy for the fourth follow-up differed from prior sweeps. The sampling strategy only included individuals that have participated in at least one NCDS sweep previously and actively excluded those known to have emigrated or to have died – there was also no attempt to include new immigrants as there was in the first three follow-ups.

Table 1.2 Sweeps Included in Analysis NCDS

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year | 1958 | 1965 | 1969 | 1974 | 1981 |
| Sweep Number | 0 | 1 | 2 | 3 | 4 |
| Sample Size | 17415 | 15425 | 15337 | 14654 | 12357 |
| Age | Birth | 7 | 11 | 16 | 23 |

#### Sample Size and Attrition

Table 2.1 details the sample size of the NCDS. At birth in 1958, the total cohort consisted of a sample 17,638 with 17,415 participants. By 1974, age 16, the total cohort had increased to 18,558. This is because the original sample was supplemented by migrants born in 1958. The number of participants at age 16 had fallen to 14,654, or 91.6 per cent of the eligible sample. This is a reduction in actual participants from the birth wave of 2,761. Of this reduction, 873 people died, and a further 799 emigrated, leaving 1,089 missing for reasons other than death or emigration. By 1981, at age 23, the total cohort was 18,558. After considering 960 dead and 1,196 emigrants, the eligible sample is 16,402. There were 12,357 participants, or 75.3 per cent of the sample.

Table 1.3 Participation in the NCDS from birth to 23 years

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Total cohort | Dead | Emigrants | Eligible sample | Participants | (% of the eligible sample) |
| Birth – 1958 | 17638 | 0 | 0 | 17638 | 17415 | 98.7 |
| Age 7 – 1965 | 18016a | 821 | 475 | 16720 | 15425 | 92.3 |
| Age 11 – 1969 | 18287a | 840 | 701 | 16746 | 15337 | 91.6 |
| Age 16 – 1974 | 18558a | 873 | 799 | 16886 | 14654 | 86.8 |
| Age 23 – 1981 | 18558 | 960 | 1196 | 16402 | 12357 | 75.3 |

The original sample was supplemented by migrants born in 1958.

The sample size at age 23 is substantially smaller than that of the initial sample; this sample attrition is primarily determined to be caused by not being able to trace participants (there is also a relatively low refusal rate – 7.1 per cent at age 23) (Power and Elliott 2006). For refusal rates, it is best to understand the dynamics of how the NCDS survey was conducted to appreciate the possibilities related to refusal. At age 23, this was the first time that participants filled out the survey independently without parental or guardian assistance (like at age 16) or having it done for them by their parents, teachers, and medical professionals. Data were collected by a paper and pencil-based survey.

Hawkes and Plewis (Hawkes and Plewis, 2006) demonstrate that ‘non-response: others’ (cases where there is no data for this sweep but there is for later sweeps, and ‘temporary emigrants’) make up 10.7 per cent of non-response. Around 6.3 per cent of the remaining missingness is categorised as ‘eligibility unknown’ (ibid). Eligibility unknown relates to those who either died or permanently emigrated. There is, overall, a substantive amount of missingness within the data used for analysis. An issue with sample attrition for the NCDS is that the size of specific ethnic minority populations when the sample was first collected was small, meaning that attrition analysing ethnic minority populations is challenging (ibid).

### British Cohort Survey

This thesis will also use data from the British Cohort Study using (Bynner, 2023; Bynner, Butler and University College London, 2023; Chamberlain, Chamberlain and University of London, 2023; University College London, 2023; University of London, 2023a; University of London, Butler and Bynner, 2023; University of London and Bynner, 2024)[[3]](#footnote-3). The BCS began in 1970 with data initially collected on 17,198 babies born in England, Scotland, Wales, and Northern Ireland in the week of 5-11th April and uses a systematic sampling strategy. This chapter will use data from participants up to the age of 30. Full cohort sweeps were gathered when participants were aged 5, 10, 16, 26, and 30 with a 10 per cent subsample taken at 21 (Bynner, 2017). Initially, the data was collected using medical records and the mother’s input. As the cohort members aged, they started actively answering survey questions. Age 26 was the first time the cohort member took direct control of answering the survey itself. This was also a period of transition for the BCS; it typically relied on school records to keep in contact with its cohort members through their registered addresses, but after the age 16 sweep, when most left mandatory education, a large number of respondents were lost when it came time to contact them for the age 26 sweep (Elliott and Shepherd, 2006).

For a full breakdown of information sources for each wave of the BCS included in subsequent analysis, refer to table 1.4. The methods for collecting information for the BCS included face-to-face interviews, proxy interviews, telephone interviews, self-complete questionnaires, assessments, and medical measurements.

Table 1.4 Follow-ups and information sources for BCS

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Year | 1970 | 1975 | 1980 | 1986 | 1991 | 1996 | 2000 |
| Age | Birth | 5 | 10 | 16 | 21 | 26 | 30 |
|  | Mother | Parents | Parents | Parents | Subject | Subject | Subject |
|  | Medical | Medical | Medical | Medical |  |  |  |
|  |  | Tests | Tests | Tests |  |  |  |
|  |  |  | School | School |  |  |  |
|  |  |  | Subject | Subject |  |  |  |
|  |  |  |  | Census |  |  |  |

Table 1.5 Sweeps Included for Analysis BCS

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Year | 1970 | 1980 | 1986 | 1991 | 1996 | 2000 |
| Sweep | 0 | 2 | 3 | (sub-sample) | 4 | 5 |
| Sample Size | 16571 | 13071 | 14874 | 1645 | 11621 | 9003 |
| Age | Birth | 10 | 16 | 21 | 26 | 30 |

Alongside the standard sweeps detailed in the table above, the 21-year-old sub-sample sweep and the BCS economic activity dataset are two other aspects of the BCS. Both were considered supplementary to the total sample sweeps. The former consisted of a sub-sample of 10 per cent of the participants in the full sample and covered aspects such as economic activity since age 16. The BCS economic activity dataset’s sole focus was creating a monthly economic activity record of participants since they left mandatory schooling up to 2016. The economic activity dataset used activity-related data from sweeps 5-10 (Hancock and Peters, 2021). Barring the 21-subsample sweep, economic activity history on the type of activity individuals did post-mandatory schooling at age 16 was not collected until participants were aged 30 in sweep 5. The content covered in the age 30 sweep and the economic activity dataset for the period of this analysis is identical. The 21 subsamples, however, provide additional data that is missing in both the economic activity and sweep six datasets. Data was thus merged with the sweep five and the subsample to boost the overall sample size of the outcome variable of interest.

The BCS sample has issues with longitudinal linkage to earlier and later datasets. The unique case identifier included with the BCS70 datasets is the 6-digit variable [KEY] derived from combining the 5-digit variable [chesno] and one-digit twin code [tc] together (Dodgeon, 2002). All participants taken at the Birth sweep were given KEYs ranging up to 200,000. Those added to the survey at age five were given KEYs from 300,010-450490, the 10-year-old sweep KEYs 600020-703560, and the 16-year-old sweep KEYs 800020-804890 (Dodgeon, 2002). KEYs were added up to age 30, but after age 16, expanding the population base was limited to returning to those already located but not already interviewed (Dodgeon, 2002). Including new participants or new KEYs at later points, post-birth sweep means that some individuals have important information missing at earlier and later sweeps within the BCS. For example, the 21 sub-sample sweep has 92.59 per cent of cases originally collected at Birth. The remaining 7.41 per cent were collected from age five onwards (ibid).

Another issue with the BCS data is that those in the original birth sample included 626 children living in Northern Ireland. After the initial survey, the Northern Ireland population was excluded from all subsequent sweeps, except for the small amount that moved to Great Britain (Dodgeon, 2002). Thus, any substantive interpretations of the dataset using data post-birth-sweep cannot draw on any Northern Irish data.

#### Sample Attrition and missingness:

Table 1.6 Participation in the BCS from Birth to 30 years

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Total cohort | Dead | Permanent Emigrants | Non-Response | Other[[4]](#footnote-4) | Participants | (% of the eligible sample)[[5]](#footnote-5) | Data Collected From |
| Birth – 1970 | 17,287 | 0 | 0 | 91 | 0 | 17,196 | 96 | Mother and Medical Records |
| Age 5 – 1975 | 16,720 | 567 | 0 | 2,812 | 360 | 12,981 | 79 | Parents, medical records, cohort members |
| Age 10 – 1980 | 16,700 | 587 | 0 | 588 | 655 | 14,870[[6]](#footnote-6) | 89 | Parents, school, tests, medical exam, cohort member |
| Age 16 – 1986 | 16,690 | 597 | 0 | 2,884 | 1,594 | 11,615 | 70 | Parents, school, tests, medical exam, cohort member |
| Age 26 – 1996[[7]](#footnote-7) | 16,545 | 697 | 45 | 4,416 | 2,384 | 9,003 | 55 | Cohort member |
| Age 30 – 2000 | 16,253 | 747 | 287 | 2,439 | 2,553 | 11,261 | 70 | Cohort member |

The BCS did not record information on permanent emigrants before sweep 4; some have attempted to estimate this number in prior sweeps (Plewis, 2004), but it remains an estimation. Another to mention is that unlike the NCDS cohort, where the number of participants has steadily declined as the sweeps go by, there is a much more tumultuous story for the BCS. The BCS went through many states of management and how the data was precisely collected; this, combined with the extensive period of 10 years from age 16 to age 26, has meant that attrition has been less than steady for the BCS cohort. This, even before analysis, suggests that missing data may present a problem for future models.

[BHPS section and UKHLS section under construction – will have more detail when datasets are known]

### The British Household Panel Survey

This analysis also uses the British Household Panel Survey (BHPS) using SN6931[[8]](#footnote-8) **[subject to change].** The BHPS was a nationally representative survey of households living in Britain. The BHPS started in 1991 and ended in 2008 – collecting a total of 18 waves, or years, of data. The BHPS collected data on 5,000 households totalling around 10,000 individual adult interviews (Taylor *et al.*, 2018). All individuals were re-interviewed in successive waves, and if split off from their original households, all adult members of their new households would be interviewed. The young were interviewed in the BHPS through a ‘Youth Panel’ that interviewed children aged 11-15 in each household from Wave Four onwards (1994). The Youth panel allows the linkage of children’s unique personal identifiers to their parents through the household identifier. This allows the linkage of parental variables to the children via the household. The BHPS used a sample of 8,167 original households drawn from the Postcode Address File (ibid). Through its random sample of Great Britain, excluding the Scottish Highlands and Islands, it achieved a sample of 5,500 households. New eligibility into the sample was contingent on either being born into the sample from an original sample member, an original sample member moving a household into a house with one or more new people or one or more people moving into the household with an original sample member. The BHPS offers three sub-sample boosts. The first is the United Kingdom European Community Household Panel (ECHP) sub-sample that provides a ‘low-income’ sample. The second is the Scotland and Wales extension samples that provide a boost to those households living in either Scotland or Wales. The third is the Northern Ireland Household Panel Survey boost, doing the same but for Northern Ireland. Each household was provided with a questionnaire package to complete, this included a household coversheet containing all household observations, a household composition form that was completed on meeting with the household by the interviewer, a short household questionnaire completed by the household reference person, and then individual schedule questionnaires completed by all adults in the household, as well as a quick self-completion questionnaire. Individuals that were missed were also followed up through a proxy schedule and a telephone questionnaire. The BHPS collected detailed demographic data for each individual and household and collected educational data for youth once they entered the adult survey for the first time.

### United Kingdom Household Panel Survey (Understanding Society)

The United Kingdom Household Panel Survey (UKHLS) builds upon the BHPS, incorporating all members into the UKHLS study. The UKHLS started in January 2009 and is an ongoing household panel survey that collects waves of data every year. The current Wave is Wave 13 which started data collection in January of 2023. The UKHLS targets 40,000 households across the study population. Each adult individual is interviewed in the adult questionnaire followed by all children in the child questionnaire. The UKHLS provides a rising 16 youth variables allowing the construction of cohorts that follow children into adulthood. The UKHLS has four sampling components: the general population sample, the ethnic minority boost sample, the Innovation Panel, and the sample of participants from the BHPS (Buck and McFall, 2011; Stopforth, 2020). Each sample design relies on multi-stage stratification and clustered samples. The General population sampling component is a stratified, clustered, equal probability sample of residential addresses drawn to a uniform design throughout the whole of the UK (the Northern Ireland sample is not clustered) (ibid). The primary sampling units are postal sectors stratified by nine regions of England plus Scotland and Wales (ibid). The overall sample targeted 40,000 households: 26,000 from the general population sample, 4,000 from the ethnic minority boost, 1,500 from the Innovation Panel, and 8,400 from the BHPS participants at Wave 18 of the study (ibid). The total achieved was 39,802 households containing 101,086 individuals. Compared to the BHPS the UKHLS does not collect educational data in the first six waves of the survey due to it being linked with the National Pupil Database (NPD). Educational variables do exist post-Wave 6 and data for previous waves can be constructed using this post-Wave 6 data.

Four-digit Standard Occupational Classification codes needed to construct social stratification measures are restricted to special license data for the UKHLS. This data is accessible using the Special Licence data accessed by the UK Data Service.

#### Complex Survey Design

Complex surveys are defined by their data collection strategies, any data that is collected by means other than a simple random sampling strategy can be considered to be a complex survey. The NCDS and BCS have a systematic sampling design and will be discussed first. Following this the stratified sampling design of the BHPS and UKHLS will be discussed. The systematic sampling strategy selects a random sample with a fixed periodic interval that is selected form a larger population. In the case of both the NCDS and the BCS the systematic sampling that follows uses a select period to allow for sample selection for all individuals born within that period.

The sampling strategies that both the BHPS and UKHLS employed make a straightforward analysis ill-advised. It is vital to reflect on the design and sampling strategies of both household panel surveys to make appropriate inferences from the sample. Whilst the BHPS and UKHLS both are complex survey designs; their designs are slightly different in construction. The BHPS has a two-stage stratified sample design with the first stage of identifying primary sampling units of postcodes followed by a systematic sampling used to collect addresses for interview (Taylor *et al.*, 2018). The UKHLS found the primary sampling units from the initial stratified sample, these sampling units were then sampled systematically with equal probability within each strata (Buck and McFall, 2011). The analysis using the UKHLS will use both the general population sample (main sample survey) and the ethnic minority boost sample – a sub-sample boost that over-samples ethnic minority populations allowing for appropriate inferences to be made about ethnic minority individuals within the overall sample.

Survey weights are provided by both the BHPS and the UKHLS to handle non-response and deal with selection of the household panel surveys. Advice on which weights to use and why is provided by Lynn and Kaminska (2010). Complex survey design packages within Stata, specifically the ‘svy’ package, will be used to make suitable adjustments for complex designs.

## Methods

All research questions in this thesis are answered using quantitative methods. The subsequent work uses large-scale, complex datasets. This thesis will use regression models from the generalised linear model (glm) as appropriate from the outcome. The following work on each cohort is broken down into two substantive sections covering: a simplistic logistic regression model analysing youths' first major transition, a multinominal logistic regression model analysing youths' destinations following their first major transition.

Part 1 will focus on an initial combined cohort analysis of the youth’s decision to continue schooling versus not continuing schooling. This is followed by a more granular analysis of each cohort by conducting a sensitivity analysis of social stratification measures and Standard Occupation Classification codes, followed by a handling missing data section. A similar procedure will be implemented for Part 2 which will use a multinominal logistic regression analysing youths’ destinations in economic activity following their first transition. Each model provided will contain log odds, average marginal effects, and quasi-variance statistics that were appropriate. Every model will also be graphed visually using predicted probabilities and coefficient plots compared with quasi-variance statistics. The methods for each set of analyses will be outlined in greater detail within each cohort section of the analysis.

### Logistic Regression Models

Part 1 will use logistic regression models, and these will be outlined now. The logistic regression model is a non-parametric, non-linear model that manages a binary categorical dependent variable (Gayle and Lambert, 2009; Uberti, 2022). The dependent variable for all models in Part 1 will be a binary categorical variable of economic activity defined through ‘continuing schooling’ or ‘not continuing schooling’. The logistic regression model requires a reference category, for all analytical models in Part 1 the reference category refers to ‘continuing schooling’.

### Multinominal Logistic Regression Models

Part 2 will use multinominal logistic regression models, and as such, these models will be outlined here. The multinomial logistic regression model is a non-parametric, non-linear model that is an extension of the logistic regression model that manages a nominal categorical dependent variable (Wulff, 2015). The multinomial logistic regression model is appropriate for dependent categorical variables with more than two categories. The dependent variable for all models in this thesis will be the economic activity of individuals after leaving mandatory education. As such all models in this thesis will contain a dependent variable with multiple categorical outcomes. The multinomial logistic regression model works very similarly to the logistic regression model, but because there are more than two categories, more calculations are required to produce the relevant statistics. Though both types of models share the need for a reference category.

### Goodness-of-fit statistics

Where relevant across all Parts, models of analysis will be compared and assessed using goodness-of-fit statistics formed of the Akaike Information Criterion (AIC), the Bayesian Information Criterion, and statistics.

This thesis will engage in multiple forms of sensitivity analysis requires a suitable measure of model selection. Both the AIC and BIC statistics offer a suitable solution to aid in model selection. The AIC is calculated from the number of independent variables that are within a given model in addition to the maximum likelihood estimate of the model (Akaike, 1998) – the model that should be preferred after comparison is the one that explains the greatest amount of variation using the fewest possible independent variables (Cavanaugh and Neath, 2019). An AIC statistic only has value in a nested context, it means nothing on its own, it only works as a comparative measure. The lower the AIC the better the model fit. The BIC statistic follows very similarly from the AIC statistic, however the BIC penalizes an increase of parameters in a given model (Neath and Cavanaugh, 2012; Profillidis and Botzoris, 2019). Both the AIC and BIC statistics will be reported alongside each other for each model presented.

There are several pseudo statistics to choose from, none appear to have a consensus on which is best or most appropriate to use (Allison, 2013). Previous empirical work on the different measures of statistics have demonstrated that for the same model, different measures produce wildly different pseudo (Smith and McKenna, 2013). Four common pseudo that are used are: McFadden’s and adjusted (McFadden, 1972), the Nagelkerke (Nagelkerke, 1991), as well as the Cox-Snell (Cox and Snell, 1989) amongst others. For a linear model, the statistic represents the proportion of variance in the dependent variable that can be explained by the independent variables in an ordinary least squares regression model. An of 0.4 in this regard would represent 40 per cent of the variance being explained. is defined as:

For non-parametric models the becomes slightly more difficult to interpret, for logistic based regression, the estimator is maximising the likelihood function. There is no ‘true’ measure of in a non-linear model, though the proportion of unaccounted-for a variance that is reduced by adding variables to the model is the same as the proportion of variance accounted for, or . All four pseudo statistics use this general logic to construct their own variations of . The interpretation of a pseudo differs from its linear regression counterpart due to the limits placed upon a logistic or multinominal pseudo- based measure. Whilst the pseduo shares with the the rule that as the limit tends to increase as the absolute value of increases with other parameters that are fixed. There is a difference in the proportion that these limits increase by, with pseudo measures increasing by a lower rate than linear counterparts, even when the associations are strong (Hu, Shao and Palta, 2006). Four pseudo are presented below.

McFadden’s is defined as:

Where is the value of the likelihood function for a model with zero predictors and is the likelihood of the model being estimated. The is analogous to the residual sum of squares in an OLS regression - analogous to .

McFadden’s adjusted is defined as:

Where K is the number of estimated parameters in the model. The adjusted version of McFadden’s penalises the as more paramters are added to the model, making it an attractive option to use.

The Cox-Snell (also known as the maximum likelihood ) is calculated as:

Where n is the sample size and represents the negative likelihood ratio chi-square statistic and N the total number of observations. The Cox-Snell can be calculated for both linear and non-linear models – the equation is identical. As Allison states, this is more appropriately termed a ‘generalised’ rather than ‘pseudo’ because the usual used in linear regression depends on the likelihoods for the models without predictors by this formula (Allison, 2013). The Cox-Snell is very attractive as it is consistent with linear measures is consistent with maximum likelihood as an estimation method, is asymptotically independent of the sample size n, and has an interpretation of explained variation (Nagelkerke, 1991).

The major issue with the Cox-Snell however is that it has an upper bound of less than 1.0 and is dependent on the margin proportion of cases within events – this means that the upper bound of a given model can be a lot less than 1.0 or very close to it, depending on the marginal proportion of cases within events. This makes the Cox-Snell much less attractive.

A solution to this presented by Nagelkerke, that is to divide the by its upper bound. The Nagelkerke (also klnown as the Craig and Uhler ) is defined as:

This ‘solution’ is ad hoc however. These tend to be the highest out of all pseudo methods.

Each of these measures present certain issues. Following the advice from Allsion (Allison, 2013) the Tjur measure, or as Tjur calls it, the coefficient of discrimination (Tjur, 2009) appears to be the best measure for use of interpretation in logistic regression models. The Tjur measure is defined as:

The Tjur measure has an upper bound limit of 1.0 and is very similar to the linear estimation, as it is calculated for each category of the dependent variable, calculated the mean of the predicted probabilities of an event, then take the difference between the two means (Allison, 2013). The Tjur is equal to the arithmetic mean of two formulas based on squared residuals and equal to the geometric mean of two other formulas based on squares residuals (Allison, 2013). Whilst there is no automatic output for this measure in Stata, it can be accomplished by after running a regression, running the predict command on an e(sample) and then getting the difference in means from a ttest.

The Tjur is not linked to the likelihood function and as a result adding additional variables to the model could result in a decline in the overall . This is a benefit rather than a detriment to the measure. This allows for a better comparative of predictive potential for model building. A major issue with the Tjur is that it can’t be readily applied to an ordinal or multinominal logistic regression. As such it cannot be readily used within Part 2 of this thesis but can in Part 1. The Tjur measure appears the most attractive, it is as of writing not applicable to a multinominal logistic regression model that is planned in this thesis, however for logistic regression models it will be reported.

Returning to the other pseudo measures, all have issues and so choosing one to go forward is a difficult task. Therefore all measures spoken of thus far will be included in subsequent model statistics – though for the sake of brevity only the McFadden’s Adjusted will be directly reported in Part 2, other measures will be reported in the tables of statistics only.

Whilst Stata does provide certain commands such as ‘estat’ that provide some related measures, and other custom commands such as ‘fitstat’ (Scott and Freese, 2001) provide even more measures, there is no single command in Stata that allows the production of all pseudo for a multinomial logistic regression in one place. Thus, a program was created to manually calculate each individual pseudo measure and display each measure all in one place.

### Nested Models and Fixed Variance

Each proposed model within this thesis will use a nested model approach to present results. Initial nested models will provide detailed goodness-of-fit statistics including the pseudo , AIC, BIC, deviance, and change in deviance either from the null if reporting singular added variables or deviance from previous model if reported additive nested models. The inclusion of comparing nested models with the same sample is to investigate and compare the size of an effect of one variable compared to the entire model – or to other added variables (Connelly, Gayle and Lambert, 2016).

This comparison of nested models is simple when analysing linear outcomes (ibid). However, in a non-linear context, interpretation is more difficult due to the fixed variance problem. Adding additional variables to a non-linear model alters the estimated coefficients within that model even when the explanatory variables are not related to one another. The size of coefficients in non-linear models may alter based on the rescaling of the model rather than any relation to other variables such as the case in a linear model context (ibid). Given this, it is important to not treat a comparison of non-linear nested models using the same sample naively. Following the advice provided by Connelly, Gayle, and Lambert (2016b), the comparison of nested models will also include a table using the Karlson-Holm-Breen (KHB) method. The KHB method estimates the changes in the coefficients in a non-linear model that are the result of rescaling when new variables are added – this can then decomposing changes in effect sizes into direct effects and indirect effects (Karlson and Holm, 2011; Kohler, Karlson and Holm, 2011; Karlson, Holm and Breen, 2012; Breen, Karlson and Holm, 2013). This analysis will provide evidence for or against the view that a reduction in log odds from a variable of interest is the result of additional variables or not. One table will be constructed for each analytical model used alongside traditional model building statistics tables. It will contain each additive nested model of interest, with each added variable the KHB method will be used and produce for each variable a reduced, full, and difference log odds statistics. These will be interpreted to understand if rescaling or additional variables is the cause of a reduction in log odds for each variable. Secondly the confounding ratio, percentage, and rescaling factor will also be reported for each nested model. The cofounding ratio indicates the total effect size and if that is larger or smaller than the direct effect. The confounding percentage indicates the percentage of total effects of a given variable is due to the additional explanatory variables added to the model. Finally the rescaling factor provides an estimated size of the total effects that are the result in rescaling the model.

## Structure of Thesis

This thesis is structured into three parts. Part 1 presents an analysis of the relationship between structural inequalities and economic activity for the National Childhood Development Study, British Cohort Study and the United Kingdom Household Panel Survey. Part 1 will analyse the impact structural inequalities have upon an individual’s first major transition – continuing schooling or not continuing schooling. The main statistical method used is a logistic regression model. All three cohorts will be combined into one dataset, and interaction effects will be used to analyse cohort-level impacts. After initial modelling, each cohort will be studied separately, with sensitivity analyses of social stratification measures and Standard Occupation Codes conducted. A handling missing data simulation section precedes any handling missing data methods implemented for each cohort. Following this, each cohort’s complete records analysis is compared to a multiple imputed model to assess the possibility of a missing at random mechanism. Where a missing completely at random mechanism is present the complete records analysis is selected. Part 1 ends with a discussion of substantive findings between and within cohorts.

Part 2 will use multinomial logistic regression models to provide a more granular assessment of the impact of structural inequalities on individual behaviour by focusing on the destination’s youth sort into post-mandatory schooling. The main statistical method used is a multinomial logistic regression. A sensitivity analysis for the cohort comparing social stratification measures and Standard Occupation Classification codes is conducted. This comparison will once more use multinominal logistic regression. A handling missing data section then conducts multiple imputations and comparisons with complete records analysis. Given that the sample for Part 2 is identical to that of Part 1, if handling missing data procedures in Part 1 determines a missing completely at random mechanism is present over a missing at random mechanism, then no multiple imputation is required. Part 1 ends with a conclusion of statistical and substantive findings.

Part 3 is a comparative section that brings together all the statistical and substantive findings across each cohort to understand the temporal trends that structural inequality has on individuals’ school-to-work transition.

# Youths First Major Transition Post-Mandatory Schooling

## Introduction

The transition from school-to-work is for most young people, the first major transition and one of the first significant life choices a young person in the UK must make. This first transition forms a bridge between the life domain of education and preparation for the world of work, and the phase of active employment (Mayer, 2009).

Much of the literature surrounding a young person’s first major transition involves the notion of an individual’s ‘choice’. Choice as a concept is a complicated affair prominent in youth transition literature (Micklewright, 1989; Schoon, 2010). The role of structural inequalities can influence individuals' choices across their life course. These structural inequalities may provide opportunities or hinder individuals' decision-making when sorting into their economic activity post-mandatory schooling. The influence of structure upon choice is dependent and influenced by the socio-historical context in which the choice is made. To analyse the role of structural inequalities on youth’s first transition as well as the socio-historical context, several datasets will be employed within Part 2 to assess trends within and between cohorts. The cohort of individuals analysed in this chapter comes from the National Childhood Development Survey (NCDS), the British Cohort Study (BCS), and the British Household Panel Survey (BHPS) and United Kingdom Household Panel Survey (UKHLS). Each cohort will be analysed and explained in detail prior to analysis. In addition, detailed literature reviews of the socio-historical contexts of each cohort will also be provided.

The cohorts from each respective dataset provide an ability to study the influences structural inequalities have on youth transitions within a socio-historical context. The following chapter identifies a simple model that maps the first transitionary pathway following post-mandatory education. This model uses a simply binary dependent variable of economic activity post-mandatory schooling: if an individual continues schooling or does not continue schooling. Figure 2.1 illustrates the first major transitional pathway possible for the individuals within each cohort, with statistics detailing the number of individuals from each cohort continuing schooling and those that do not continue schooling.

There are 12 years between the birth of the NCDS cohort and the birth of the BCS cohort. Within 12 years the proportion of individuals continuing schooling flips from a minority to majority. The UKHLS cohorts all see an increased per centage of continuing schooling compared to the NCDS and BCS cohorts. This is a sociologically compelling reason to study these cohorts together, to uncover the within as well as the between effects to uncover the socio-historical contexts that govern and influence individuals sorting and transitionary experience. There appears to be a steady positive monotonic increase in individuals continuing education post-mandatory schooling. Whilst there is a clear descriptive picture that emerges from figure 2.1, further investigation is required to understand the role of structural inequalities and their potentially changing influence on youths first major transition.

A close-up of a black and white sign

Description automatically generated

Figure ‎2.1 The Story of First Transitions by Cohort [Re-do when UKHLS is integrated]

## Literature Review: Cohorts in Context

Each cohort will now be discussed with relevant empirical literature. This literature review will ground the subsequent analysis in existing analysis of youth transitions as well as presenting limitations in older literature that can and will be improved upon. The review will start with the NCDS cohort, a time of post-war economic re-structuring with implemented educational reforms changing the educational landscape. Following this, the BCS cohort literature focuses upon a declining manufacturing sector, a collapsing youth labour market, and reform to apprenticeship and training programs. Finally, the UKHLS cohort literature provides context for further educational reforms, political change, and another re-formulation of training and apprenticeship programs. This literature review not only provides context for each cohort used throughout this thesis, but it also provides a reference point to understand statistical phenomena – and with the adoption of contemporary statistical techniques, to verify prior literature on the topic of youth transitions for each cohort. The primary finding across all cohorts presents compelling evidence that structural inequalities have a continued impact upon young people’s transitionary pathways. This provides a contextual justification for subsequent analysis.

### NCDS in Context

Trajectories into further education, apprenticeships, employment, and unemployment are identified within the NCDS cohorts’ socio-historical context. Each of these trajectories appears to have an element of structural inequalities influencing the outcomes of individuals. Within the timeframe of the NCDS, much focus is placed on the role of social class and sex.

The evidence provided through empirical literature is also expanded upon by reviewing the theoretical and historical literature to provide a holistic overview of the school-to-work transition during the NCDS period.

**Compulsory Schooling**

Young people were in full-time compulsory education until they were 16. The NCDS youth were part of a larger cohort of children impacted by the rising school-leaving age (ROSLA) in 1972. At 16, individuals were typically expected to sit some form of examination. This was a mixture of Certificate of Secondary Education (CSE) (Pearson qualifications, 2023a) and Ordinary level (O’level) (Pearson qualifications, 2023b). The NCDS cohort were some of the first young people to experience a reformed education system post-World War Two. This reformed system would provide a bedrock foundation for all future reforms.

#### Story of transitions for NCDS youth

From the age of 16, NCDS youth had multiple pathways. One popular pathway was to enter the labour market and employment straight away. Another growing pathway continued education, thereby elongating their educational pathway. Traditionally, this would mean joining a sixth-form college and taking Advanced levels (A’levels). During the NCDS period, non-traditional[[9]](#footnote-9) educational pathways were also available that provided greater possibilities to stay withing education even if a young person was not particularly academically gifted. These non-traditional pathways typically led young people to technical colleges offering a set of non-traditional qualifications. Beyond education, there was also the option of joining a training & apprenticeship program. These were mainly geared towards specialised manufacturing labour. The apprenticeship programs of the NCDS cohort provided highly skilled training, typically in a manual field, which almost always guaranteed a job at the end of the programme. Finally, some would enter a period of unemployment or opt to be out of the labour force. The latter of these pathways was particularly prominent for women during this period, especially after pregnancy and/or marriage.

**Youth Labour Market for NCDS youth**

The latter half of the 20th century was plagued by exogenous economic shocks, demand for economic restructuring, and a changing economic landscape. These circumstances prompted many committees, programmes, and reports on the topics of assessing and providing solutions to Britain’s lagging training, education, and economic performance. As far back as 1942 a joint consultative committee was formed to assess the problems of the technical education and industrial training of youth (Deakin 1996). The observations made from the committee were that pre-employment vocational training was not developed to the same extent as in other countries and systems of apprenticeships were not utilised to their fullest potential (Parker, 1957). This sets the theme for the next 60 years of vocational training in the UK, whereby training programs within the UK are seen as inadequate, and comparative to other countries programs, inefficient.

The state of the British economy during the 1950s was one of tight levels of employment (Deakin, 1996). The Carr committee warned that the baby boom that would enter the labour market in the late 1960s would require a large scale increase in apprenticeship schemes to maintain British manufacturing skill (*Carr Committee Report on Recruitment and Training in Industry*, 1958). This ‘manpower’ problem was a key focal point of education and economic discourse that followed the NCDS cohort in their early years. One of the implications of the manpower situation was the establishment of industrial training boards following a White Paper on industrial training (‘Industrial Training: Government Proposals’, 1962). The Industrial Training Boards were established to combat the uncoordinated quality of existing training schemes as they were almost entirely left in the hands of individual firms within the market. Following this White Paper, the Industry Training Act was passed in parliament in 1964, which attempted to centralise training schemes for the young by setting up the Central Training Council. The Central Training Council advised Industrial Training Boards who managed a levy grant system on training - by 1969, 27 Industrial Training Boards existed (Perry, 1976). Trainees using this scheme were paid a tax-free allowance by the government in the 1970s - between eight and 14 pounds according to family circumstances. Approximately 90 per cent of trainees who completed their courses were placed in jobs in the trade for which they were trained (Orchard, 1970). Whilst successful, the scheme did face critiques, central to these was: a lack of central direction because of the advisory nature of Central Training Council, large firms continued their own internal training schemes, the training arrangements were employer based and industry orientated over workers (Deakin, 1996).

Based upon these critiques, the Industrial Training Boards system was reformed through the Employment Training Act 1973. The major reform of this Act was the establishment of the Manpower Services Commission (MSC). The shift in demand for certain skills based upon structural shifts in the British economy resulted in the establishment of the Training Opportunities Programme (TOP) in 1972. Whilst the MSC and TOP programme see a continuing centralisation of youth training in the British labour market, the first direct intervention for the British state since World War Two came in the establishment of the Temporary Employment Subsidy (TES) in order to tackle total unemployment (Deakin 1996). The TES in part was a forced measure due to lacking a headed call from the Carr Committee. Those supported by the TES represented 6.1 per cent of all employees from the manufacturing industry and 2.2 per cent of the total labour force (TES only supported private employees). The overrepresentation of manufacturing employees in comparison to the total labour force provides explanation for some of the root causes of economic uncertainty during the NCDS context. Manufacturing jobs, and the wider manufacturing industry was facing the most strain during this time – skilled manual labour was struggling.

The baby boom of the 1960s, was in the late 1970s starting to enter the labour market for the first time. The population of 16-19 year olds was 2.9 million in 1971 and by 1976 was 3.6 million - an increase of 23.5 per cent (Deakin 1996). The NCDS cohort would leave mandatory education in 1974. NCDS youth would be leaving school near the peak of the school leaver unemployment boom.

Due to this bulge in the youth cohort and a lack of policy to accommodate this increase, school leavers saw unemployment increase sixfold from 13,300 in 1974 to 81,600 in 1976 (ibid). As a result of increasing youth unemployment directly following mandatory education, the Recruitment Subsidy for School Leavers (RSSL) was established in 1975 offering employers 5 pounds per week for six months per school leaver recruited (Deakin 1996). Following a survey of participants in the RSSL scheme, 76 per cent of all employers stated that they would have recruited as many school leavers without the subsidy (ibid). The RSSL scheme was terminated in October 1976. The RSSL scheme and the Youth Employment Subsidy (YES) that followed it were both subsidies aimed at the firm or employers, rather than at young people. The YES scheme lasted from 1976-1978 at a cost of 8.7 million pounds and supported 38,970 places in firms. Similar to the failures of the RSSL, a survey conducted for the YES scheme indicated 75 per cent of firms would have employed the same number of young leavers without the scheme (ibid).

It wasn't until the Work Experience Program (WEP), established in 1976 that the government attempted to directly subsidise the employee rather than employer. The WEP provided subsidised work placements for young people aged 16-18 by providing them with 18 pounds per week which was not eligible for income tax or national insurance contributions (ibid). A total of 61 per cent of individuals on the WEP scheme attained a full-time job after leaving (Lasko, 1978).

Whilst programmes like the WEP did have some impact on providing young people with adequate post-schooling opportunities, other programmes such as the RSSL and YES were failures. For this reason, youth unemployment remained high into the late 1970s. In 1977 youth unemployment of school leavers was at 99,000 representing 7 per cent of total unemployment (Deakin 1996). The Youth Opportunities Programme (YOP) was established as one of the first large scale youth training programmes in 1978. Over the course of YOPs existence, 1,834,700 million people accessed the programme. The YOP offered two types of schemes: work experience placements of up to six months in private firms, and work preparation places that offered 13-week remedial education. It's lack of long-term training placements meant that in 1983 it was replaced by the Youth Training Scheme (YTS).

The youth labour market of the NCDS cohort can be characterised as a declining potential pathway for certain young people. The abject decline of skilled manual occupations through the collapse of heavy manufacturing in conjunction with the demographic baby boom resulted in a lack of apprenticeship programs provided to young people – as well as access to secure job opportunities. The state attempts to alter this fate through public policy initiatives proved ineffective resulting in high levels of unemployment post-mandatory schooling. These changes in the labour market directly impacted individuals that would originally plan to enter into apprenticeship programs or enter employment in skilled manual occupations – though a much wider impact across employment opportunities was felt due to the demographic baby boom. The focused impact of this labour market change provides incentive to study the structural impacts of youth transitions – the NCDS cohort appears to have a particular historical linkage of a decline in skilled labour jobs and a rise in population. This two factors combined have serious implications for transitionary pathways for certain sub-groups of the population – namely children of skilled manual occupations, and men.

**Types of transitions the NCDS cohort experienced**

Some youth transition theorists argue that in the mid-20th century, transitions were smoother and more simplified (Vickerstaff, 2003; Leuze, 2010). During the time of the NCDS, the range of choice has been argued to have been narrower compared to later cohorts, thus owing to a more homogenised pathway (Goodwin and O’Connor, 2005). It is accurate that the NCDS cohort had large homogenous clusters related to transitioning out of mandatory education. However, the delineation between school and employment is not a strict binary – with many youth engaging in the youth labour market whilst still in education (Bynner, 2012). Youth’s choices on what they wish to do after mandatory education are influenced by labour market restructuring and recession and other factors such as the child's family background, parents, teachers, siblings, and contemporaries. While the pathways the NCDS cohort may have been able to choose from were narrow due to their socio-historical context constraining ‘choice’, this is not synonymous with smooth or straightforward. For example, individuals may have faced a seemingly homogenous experience after a period of schooling ending up in a period of employment, but that does not indicate the relative smoothness of getting from A to B. Some individuals may have experienced periods of unemployment during this time, others jumping from job to job, etc. The school-to-work transition for NCDS youth is complex, and an appreciation of choice and opportunity – and the subsequent restriction of choice and opportunity based upon structural factors need to be considered.

Martin et al (2008) used optimal matching analysis to identify the major ‘clusters’ of economic activity that individuals from the NCDS cohort enter into post-mandatory schooling. A cohort that exhibits complex, heterogeneous transitions would have many clusters that were evenly spread out. Results from Martin et al. (2008), however, convey the NCDS cohort has experienced homogenised pathways; these results are affirmed by Goodwin and O’Connor (Goodwin and O’Connor, 2005). Comparatively, Martin et al. (2008) finds that of the NCDS cohort, 96 per cent of men could be grouped into six of the most significant transition ‘clusters’ compared to 90 per cent for the 1970 British Cohort Study.

For the NCDS, the predominant pattern was to leave school post-16 and move directly to employment (Schoon, 2007). This is supported by Anders and Dorsett (2017), where transition patterns among school leavers entering the labour market were examined. They found that under the NCDS cohort, there was a large (91 per cent) number of people entering the labour market straight after mandatory schooling. This, once again, supports the view that the NCDS cohort exhibited homogenous pathways of transition. The typical pathways that young people within the NCDS entered demonstrate that a school-to-employment transition was dominant.

Training and apprenticeship programs were also a vital transition pathway – above that of continuing full-time education (Schoon *et al.*, 2001). The NCDS cohort was caught in a period of severe diminishing influence of apprenticeships. For example, the number of apprenticeships in British manufacturing declined from 240,400 in 1964 to 155,000 in 1979 (Blanchflower and Lynch, 1992). There was a severe gender bias regarding apprenticeships at this time - when the NCDS cohort was 16 years old, 40 per cent of male employees were apprenticed compared with only 8 per cent of females (Blanchflower and Lynch, 1992).

Whilst (Schoon *et al.*, 2001) found that young people from less privileged backgrounds were more likely to be in training or apprenticeships. The declining state of apprenticeships and British manufacturing has a disproportionate level of impact on young people from less privileged backgrounds. Further research suggests that apprenticeships amongst the NCDS cohort were more likely to be offered to children of fathers who were skilled manual workers over their semi-skilled counterparts (Booth and Satchell, 1994). This suggests a fragmentation of the traditional manual/non-manual divide, with a hierarchy of skills impacting the choice and opportunity of the NCDS youth.

Within the NCDS cohort, training and apprenticeships typically lead to subsequent full-time employment (Schoon *et al.*, 2001). Schoon and their colleagues (ibid) suggest that this is primarily because apprenticeships during the NCDS period spanned three years or longer, providing the relevant skills and development for young people to effectively transition from a period of apprenticeship training into stable employment. Vocational-based education is generally considered a smoother transition from school to work than academics. While this short-term benefit is worth considering, long-term disadvantages such as lower employment and wages impact those individuals with lower vocational education (Brunello and Rocco, 2017). This phenomenon has broken down post-NCDS with the breakdown of traditional apprenticeship and training programs in the UK (ibid).

Work-related training, or training on the job, has been lauded as a way for those who enter the labour market with relatively low levels of education to build up necessary skills. (Arulampalam and Booth, 1997) suggests the opposite is, in fact, the case. Work-related training seems to boost the already well-educated and leave those less educated behind. In a later study, (Arulampalam and Booth, 2001) reaffirm their findings by stating that while work-related training does improve wages, it positively affects the wages of the well-educated more so than the less-educated in the labour market. The fact that those who happen to be well-educated are related to those who come from advantaged social class positions demonstrates that advantage breeds advantage (Machin and Vignoles, 2005). Those from less affluent backgrounds who engage in work-related training will not see equal levels of growth associated with their affluent peers (Arulampalam and Booth, 2001).

The relative prevalence of employment and apprenticeship training over educational pathways suggests that the NCDS cohort experienced a pre-credentialed labour market post-mandatory schooling (Bynner, 2005). The NCDS cohort experienced a labour market that did not place challenging roadblocks to employment based on educational credentials. It was not until the 1980s that failing to get qualifications hindered getting work in Britain (Bynner, 2005).

The labour market in the pre-1980s was able to absorb people into large numbers of unskilled jobs (ibid). Those who did struggle to get jobs in the NCDS cohort were significantly more likely to experience a ‘Not in Education, Employment, or Training’ (NEET) status going forward post-21 years old (Bynner 2005: 378).

It has been established that the NCDS cohort exhibited a comparatively more homogenous transitional experience to later cohorts such as the BCS. The NCDS birth cohort did not experience a straightforward, smooth school-to-work transition. Teenagers who were still in education typically engaged in what is known as the youth labour market (Bynner, 2012). Most of this work was part-time during educational studies (Dustmann *et al.*, 1996). It is estimated that youth aged 16 worked an average of six to nine hours a week and modal earnings in the range of £1-£2 a week while still in full-time mandatory education (Dustmann *et al.*, 1996). Within the NCDS cohort, half of 16-year-olds in 1974 had a part-time job during term time (ibid). This suggests that a straightforward delineation separating school and work is an oversimplification for the time. Youth were engaging in schooling and employment before choosing what to do after mandatory schooling. Students employed during mandatory education were less likely to choose to continue education post-mandatory schooling (Neyt *et al.*, 2018). This adds a layer of complexity to the aforementioned ‘homogenised pathways. Structural inequalities – in the form of family background and unemployment status – have a role to play in the choices and opportunities of youth transitions. Homogenised pathways are not the same as smooth transitions. Structural inequality adversely impacts the relative smoothness of an individual’s transitional experience.

The relative ‘smoothness’ of youth transitions from school-to-work is primarily dependent upon the relative stability of the labour market that such individuals are transitioning into. During the time of the NCDS cohort, the labour market was experiencing a significant period of restructuring – some have also argued that the ‘collapse’ of the youth labour market also contributed to a relative amount of instability (Bynner, 2012). The notion that the youth labour market ‘collapsed’, indicating a sudden affair, is not precisely accurate. The youth labour market saw a relative decline post-war as part of broader economic restructuring. Nevertheless, the decline of the youth labour market still impacted the options available to NCDS youth.

The labour market during school-to-work transitions for the NCDS cohort was unstable and had comparatively heightened uncertainty (Leuze, 2010). The collapse of the youth labour market in the early 1980s was not a sudden affair (Bynner, 2012). Between January of 1972 and January of 1977, unemployment among 16 and 17-year-olds rose by 120 per cent (Maclure, 1978). Contributing factors such as the demise of heavy industry, the collapse of community networks, and the technological transformation of modes of production were all forces that the 1958 cohort was facing during their biographical lifespan (Bynner, 2012). The importance of the collapse of the youth labour market relates to introducing uncertainty at a critical stage of development within a young person’s life (Maclure, 1978). This uncertainty can adversely impact individuals' life domains (Mayer, 2009). These periods of instability are documented in detail with monthly employment histories (Leuze, 2010), suggesting that the collapse of the labour market impacted many individuals. This heightened instability during a time of transition for the youth of the NCDS presents an influencing factor in the role of choice and opportunity. When the labour market was facing severe restructuring, a collapsing youth labour market, and a significant economic recession, the choices and opportunities of young people seeking to transition into the world of work would be constrained and influence their choices.

**Risk and Uncertainty**

The restructuring of the economy and wider labour market during the NCDS cohorts timeframe injected an element of uncertainty and risk within the NCDS starkly contrasts the theory of ‘late modernity’ - entailing notions of risk and uncertainty in a society that provides individuals with more choice, promoting greater risk (Beck, Giddens and Lash, 1994). The literature has demonstrated that the NCDS cohort experienced comparatively homogenous transitions that were often complicated by structural inequalities that impacted the role of ‘choice’ and the individual. Structural inequalities impacted the choices and opportunities within these different pathways. The notion of ‘Late Modernity’ (Giddens *et al.*, 1991; Beck, Giddens and Lash, 1994) is based upon the idea that in the past, more concrete certainties have given way to more fluid and dynamic notions of adult identity and its development (Bynner, 1998: 31). These past certainties gave rise to stability; these current dynamics gives rise to risk (Beck, 2014). Whilst the NCDS cohort did experience relatively homogenised transitions compared to others (Martin, Schoon and Ross, 2008), the theory of Late Modernity is a naïve one that doesn’t appreciate the complexity and nuance of socio-historical context that the NCDS cohort experienced. The re-structuring of the economy and decline of heavy manufacturing industries, the collapse of the youth labour market, and the early 1980s recession are key points of risk and uncertainty that undermine the late modernity position that concrete certainties existed – it would be more appropriate to say that comparatively homogenised pathways of the NCDS cohort were smoother than those of future cohorts though risk and uncertainty remained prevalent. There is debate over how fluid certainty and choice have become; Gayle et al. (2009) provide a more updated version of events that appears to review and ultimately question the late modernity outlook.

**Educational attainment**

This theme of constraint is evidenced in the changing influence of educational attainment during the short term for the NCDS cohort. Educational attainment – and staying within education post-mandatory schooling - protects from unemployment (Bynner, Wiggins and Parsons, 1996). Those of the NCDS cohort that stayed within education post-mandatory schooling initially had higher unemployment levels due to exogenous shocks of rising national unemployment. Whilst experiencing short-term levels of unemployment, in the long run, individuals who stayed on within education had a long-term advantage in income over their peers who did not stay on within education (Payne, 1987).

Looking in more detail at educational attainment within the NCDS, individuals in the UK who choose to stay on at school post-16 were a small minority and were low by Organisation for Economic Co-operation and Development (OECD) standards compared to other Western countries (Micklewright, 1989). Compared to their non-manual peers, individuals from manual backgrounds were less likely to stay on post-16 (Micklewright, 1989). Bynner and Joshi (Bynner and Joshi, 2002), as well as Schoon (Schoon, 2007), found that young people from working-class backgrounds were less likely than middle-class peers to remain in education post-mandatory schooling.

#### Structural Barriers to successful transitions

Throughout the story of NCDS youth, a common theme of structural barriers and inequalities influencing choice and opportunity has been identified. This next section seeks to explore these structural dimensions more closely. The roles of sex, social class, and housing tenure will be explored in greater detail to provide clarity to the current empirical consensus on these forms of social stratification about NCDS youth.

##### Sex

Choice and opportunity within the school-to-work transition of the NCDS youth are influenced and impacted by structural inequality factors like sex (Dolton, Joshi and Makepeace, 2002; Makepeace, Dolton and Joshi, 2004; Cebulla and Tomaszewski, 2013). Within the NCDS, women’s roles within the labour market have marked differences from their male peers (Dex and Bukodi, 2012). Women are more likely to achieve their educational aspirations than men (Cebulla and Tomaszewski, 2013) within the NCDS cohort. They also often have higher occupational aspirations compared to men at a young age (Schoon 2007; Schoon, 2022). These aspirations however rarely translate to higher than average incomes and labour market segregation remains, whilst pay improvement for men continues to outpace women’s (Dolton, Joshi and Makepeace, 2002; Makepeace, Dolton and Joshi, 2004).

Gender segregation within the labour force for the NCDS cohort slightly declined compared to previous cohorts due to the simultaneous decline in traditionally male-dominated heavy industry labour and the growth of soft-skilled service-based employment, but overall, gender segregation remained consistently stable (Guinea-Martin and Elliott, 2008; Lekfuangfu and Lordan, 2022).

Whilst the labour market for the NCDS cohort remains somewhat segregated, social mobility does not significantly vary by gender for full-time workers (Bukodi, Goldthorpe and Kuha, 2017), though research by Savage et al (Savage and Egerton, 1997; Savage, 2011) does emphasise the impact gender has on social mobility. Part-time female workers have highly varied pathways (Connolly and Gregory, 2010). While broadly speaking, the NCDS cohort experiences homogenous transitional pathways, some sub-groups, like female part-time workers, experience a much more complex, less smooth transition into employment (Dex and Bukodi, 2012) starting from their initial higher participation in part-time work (Dustmann *et al.*, 1996).

##### Social Class

Social Class-based structural inequalities impact the educational attainment of NCDS youth during mandatory schooling (Galindo-Rueda, 2003; Sianesi, Dearden and Blundell, 2003; Holm and Jæger, 2011). This then consequently has an impact on transition outcomes and later life chances.

Evidence suggests that those individuals with advantaged social class family positions see occupational earnings increase by at least 7 per cent compared to disadvantaged family positions (Connolly, Micklewright and Nickell, 1992). Conversely, three or more months of unemployment is associated with a fall in occupational earnings by around 7 per cent (ibid). Unemployment at the youth stage increases the likelihood of unemployment at the adult stage of the life course (Gregg, 2001). To summarize, more disadvantaged groups are more likely to experience youth unemployment.

##### Educational Attainment and training

When looking at educational attainment, it is essential to consider that family background variables like social class matter (Machin and Vignoles, 2005), with the most advantaged youth seeing the best income returns (Sianesi, Dearden and Blundell, 2003). Variables such as parental education play a more critical role in the life chances of young people than parental income (Feinstein, Duckworth and Sabates, 2004; Field, 2010). Early success in education confers an advantage in later educational attainment and labour market experience (Dolton, Makepeace and Marcenaro‐Gutierrez, 2005). Educational attainment leads to more educational attainment. Achieving while young impacts educational attainment at later parts of the life course (Hutchison, Prosser and Wedge, 1979). As such, the influence of family background on early educational attainment appears to influence later life chances. Whilst educational inequality has declined in the NCDS cohort compared to younger cohorts such as the BCS (Blanden and Macmillan, 2014), it persists when translating educational attainment into the most successful occupational outcomes – those from privileged backgrounds are more likely to gain access to the highest-paying occupations, leveraging their educational qualifications.

Educational attainment translates to higher levels of income in later life—individuals with higher educational ability experience faster wage growth than their lower-ability peers (Galindo-Rueda, 2003). Early successful educational attainment is influenced, however, by a structural class effect. Those from working-class backgrounds are less likely to succeed in terms of educational attainment than their non-working-class peers (Machin and Vignoles, 2005). Some argue that this is due to poorer families being less likely to invest in education over their more affluent peers (Chevalier and Lanot, 2001). However, the nature of what constitutes ‘investment’ in an individual’s education is left unclear and subject to speculation.

Low levels of qualifications and educational attainment are related to higher propensities toward unemployment (Bynner and Parsons, 2000). The propensity toward experiencing unemployment also has a social class effect, with the growth in unemployment during the 1970s being attributed to the subsequent decline in the manufacturing sector linked to working-class labour (Schoon *et al.*, 2001). Those who are unemployed also appear to hold the lowest levels of employment commitment when they eventually enter employment (ibid). Unemployment is found within the NCDS cohort to have a scarring effect on potential earnings (Gregg, 2001; Bynner, 2012; Schoon, 2020) – the youth labour market thus plays a vital role in establishing adult future earnings (Gregg, 2001). Data suggests that a scar from early unemployment can have an estimated 12-15 per cent damaging impact on income at age 42 (Gregg and Tominey, 2005). The influencing impact that social class has on educational attainment and propensity toward unemployment appear to have long-term consequences for later life chances. The impacts of social class on youth transitions from school to work are felt in the short and long term.

The returns of higher education degrees – in other words, the income gained from educational attainment - appear to be substantial within the NCDS cohort. These returns, whilst generally lower than undergraduate degrees, also exist for higher degrees and non-degree higher education courses (Blundell *et al.*, 2000; Blundell, Dearden and Sianesi, 2001). Another study found that each successive qualification level at the National Vocational Qualification classification corresponds to a 5 per cent rise in income (Conlon, 2001).

### BCS in Context

This section provides an overview of the literature within the field of youth transitions of the BCS cohort. The structure of this section emulates the NCDS in Context section. Initially, the literature will focus on the historical and temporal context of the BCS cohort to ground the empirical research on transitions. Major transition themes are identified as they relate to employment, education, training, and unemployment. Each is influenced by structural factors that impact individual choice and opportunity. The changing nature of the labour market and British polity during the BCS period have had a substantive impact on the role of training and apprenticeships within an individual’s first significant transition from mandatory schooling into the world of economic activity.

Young people were in full-time education until 16 – like the NCDS cohort. At this age, individuals were typically expected to undergo some examination. The BCS cohort were some of the last individuals to sit the O’level at 16 before its replacement with the GCSE. After this mandatory schooling period, there were options of continuing within education, moving on to training under the YTS scheme, entering employment, or becoming unemployed or out of the labour force. The relative diversity of options compared to the NCDS cohort was restricted. Traditional apprenticeship schemes were gone, as was unemployment benefit for individuals aged 16-18. These effects will be discussed at length in the literature review below.

#### Story of transitions for BCS youth

Within Britain, the 1970s and 1980s were periods of large-scale transformation (Bynner, Ferri and Shepherd, 2019). The 1970 BCS cohort can be characterised by a continuing decline in manufacturing and apprenticeships, high levels of unemployment, more significant government intervention in young people’s economic activity, and a growing higher education participation rate following the growth of a credentialed labour market. The BCS cohort can be characterized through one word – change.

#### Youth Labour Market for BCS Cohort

The monetarist policies of the Thatcher administration since the 1979 election and following the 1980s recession fundamentally reshaped the structure of British society and the labour market. Whilst economically the Thatcher administration adopted a laissez fair policy, the state of the youth labour market prevented it from fundamentally altering course on the interventionist policies undertaken by the previous Labour government (Deakin, 1996).

The 1970s saw a continuing trend post-war, of simultaneous growth of automation and technology alongside a decline in manufacturing. However, this came more out of the 1973-5 recession that devastated the heavy industrial markets of the North of England – the recovery and rebuilding of a service economy were located exclusively within the South of England (Hamnett, McDowell and Sarre, 1989) – half of all jobs created between 1983-87 were created in the south-east (ibid). These pressures brought about the primary labour market and societal transformation for society, increasing the worker's uncertainty and risk (Schoon, 2007; Beck, 2014). As a result of this transformation of society, Hutton describes this period of British history as the ‘30/30/40’ society, whereby 40 per cent of the population are permanently in casual employment, 30 per cent are doing fine, and another 30 per cent are struggling, leading to the phrase ‘Getting on, getting by, getting nowhere’ (Bynner, Ferri and Shepherd, 2019). Whilst the baby boom demography of the NCDS promoted a general feeling of change – this was primarily located in skilled manual occupations. The widespread alterations of British society during the BCS cohort was endemic of a societal shift that impacted nearly every segment of the population.

Through this ‘transformation of society’ new jobs were created to replace the old traditionally manufacturing jobs of the past. These ‘new’ jobs were defined by their transferable skills across the service sector (Bynner and Ferri, 2003). As a result, the apprenticeship scheme linked to traditionally heavy manufacturing and highly specialised training declined. It was eventually replaced by the Youth Training Scheme in 1983 under the management of the Manpower Services Commission. The imbalance of the youth labour market and the rise of youth unemployment reached a point in 1981 whereby the MSC established national objectives following the government White Paper (‘A New Training Initiative: A Programme for Action’, 1981).The major difference between this policy compared to prior interventions was that it directly focused on youth training rather than alleviating new unemployment (Deakin 1996). Three subsequent schemes were established. The first, the Youth Training Scheme that guaranteed all young people a full year foundational training for those leaving school without jobs. The second, the Youth Worker Scheme (YWS), which aimed to improve job security and training of young people under 18 by lowering their wage rates relative to older workers. The final scheme was the Technical and Vocational Education Initiative (TVEI) to provide pupils from the age of 14 with a four-year course of full-time technical education including relevant work experience (Deakin 1996). The TVEI in particular was tied to the establishment of the National Curriculum (NC) under the Education Reform Act 1988.

The YTS was the first time in Britain that youth had become a category of large-scale policy intervention beyond education (Wallace and Cross, 1990). The YTS modus operandi was based on keeping kids off the streets and filling unemployment gaps – this became especially apparent during the recession of 1986-7, whereby the unemployment rate for men was 2.6 per cent, but 12 per cent were in some form of government training. However, this eventually fell below unemployment figures in 1988 post-recession (Bynner and Ferri, 2003). The YTS has experienced sociological critique (Droy, Goodwin and O’connor, 2019). The main critique of the YTS is that it was seen as an attempt at direct intervention from a collapsing youth labour market from an anti-interventionist government (ibid). Though other critiques focus on the YTS furthering existing divides in British society, The scheme itself appeared to marginalise women (Cockburn, 1987) and ethnic minorities (Gow, 1987) by placing them on less robust placements (Furlong *et al.*, 2017). It started as a one-year program in 1983 (eventually to a two-year program in 1986) that mainly provided low-level training that was more comparable to an alternative to unemployment than to higher education or employment (Bynner and Ferri, 2003). Whilst the YTS maintained a steady average of 400,000 people between 1985-89, it was neither an adequate replacement for the highly skilled training of a traditional apprenticeship nor an acceptable form of pay and employment. Wallace and Cross (1990) argued that the YTS represented a ‘dual-carriageway’, attempting to complete the goals of education and work training at the same time - unsuccessfully. The YTS was also internally stratified. It offered attractive, highly trained schemes, such as the so-called ‘Model A’ schemes that worked directly with employers. However, these were very hard to acquire and often went to those who did not need them the most (Wallace and Cross, 1990). The ‘Model B’ schemes were the most numerous and typically what people mean when they describe the YTS. Among these unattractive schemes, individuals were usually sorted into the growing service sector, associated with insecurity and risky employment prospects. This liminal zone of the youth labour market was stratified along gender and class grounds (Droy, Goodwin and O’connor, 2019).

It was, for many, a stopgap – an unattractive one at that. It would not be accurate to compare the YTS –a training scheme, to the much more rigorous training and education of a traditional apprenticeship (Bynner et al., 2002). Most young people felt forced into the YTS scheme because the Thatcher government cut unemployment benefits for all people between the ages of 16-18 in 1988. This is arguably the start of the punitive approach toward unemployment and welfare in the late 20th century (Droy, Goodwin and O’connor, 2019). Due to the timing of these unemployment benefit cuts, the 1970 cohort could still claim benefits. However, they still suffered as part of the ‘vulnerable core’ of the labour market through Thatcher’s cuts and de-regulations towards employment rights and the minimum wage (Hamnett, McDowell and Sarre, 1989). The proclamation in 1981 under the New Training Initiative of heralding in universal youth training for all was, in reality, a poorly thought out scheme that some compared to a stopgap, whilst harsher critiques referred to it simply as ‘slave labour’ (Bynner, Ferri and Shepherd, 2019). The YTA offered cheap, subsidised labour to employers with no requirements to continue an individual’s employment after the scheme was completed (Droy, Goodwin and O’connor, 2019). It would be fair to characterise the YTS as a short-term benefit to businesses whilst leaving the individual worker under-trained, underpaid, and often unemployed.

The initial desired purpose of the scheme was to establish a training scheme comparable to German lines (at the time, argued to be the best apprenticeship program in Europe). The result, however, was a scheme that failed to train youth appropriately, and the best form of vocational training was instead found to be employment itself (Bynner et al., 2002). The YTS has been found to have had negative consequences for men’s employment prospects (Dolton, Galinda-Rueda and Makepeace, 2004; Droy, Goodwin and O’connor, 2019; Goodwin et al., 2020) and overall a negative impact on earnings over the life course (Dolton, Galinda-Rueda and Makepeace, 2004) compared to those men who did not enter the YTS. For women, the effects on earnings were small and insignificant (Dolton, Galinda-Rueda and Makepeace, 2004).

Reflecting shortly on the YWS scheme mentioned previously. The YWS scheme was different in name only to previous subsidy schemes such as the RSSL and YES. The critiques of both prior schemes can be levelled against the YWS scheme and its successor in 1986 - the New Workers Scheme (NWS). Finally, referring back to the aforementioned TVEI scheme. The TVEI was also heavily critiqued for a number of reasons. The first critique was that whilst innovative in its approach, the release of the NC made a lot of the TVEI redundant on establishment. The second critique came from individuals that felt the TVEI was far too focused on the needs of employers rather than providing potential employees with robust adequate training for the future. The third critique was based on the fact that the TVEI was not sensitive enough to the needs of the economic landscape (Deakin, 1996). The National Association of Teachers in Further and Higher Education (NATFHE) carried multiple motions expression opposition to the TVEI through the 1980s (ibid).

From the 1980s onwards, there were three main routes from school-to-work in the UK. The first was the general academic route, that saw young people moving on to A'levels and higher education. The second route used National Vocational Qualifications (NVQs) which were occupational qualifications designed for on the job training utilising the YTS. The third route were called 'vocational A'levels'. These were obtained by staying on in school and studying for General National Vocational Qualifications (GNVQs) (Deakin 1996). Each of these three routes all require an extended or protracted stay within some form of education, be it traditional or non-traditional. Whilst the NVQs system was inextricably linked to the YTS system, and as such critiques for one can be considered critiques for the other, GNVQs were only established in 1991 and need to be discussed further. Whilst GNVQs were advertised as being 'vocational A'levels' (Deakin 1996), in reality they were nowhere near as robust in either breadth or depth. GNVQs offered a lacklustre general education and severely lacked theoretical education on mathematics and technology (Smithers, 1993).

The 1970s onwards has seen a development of falling reliance on labour and an increasing level of profit, in part related to increasing technology. Roberts called this phenomena 'jobless growth' (Roberts, Dench and Richardson, 1987). The increase in profit with a decline in jobs offered disproportionality affected the youth labour market and manufacturing industries. Though from the 1970s onwards similar trends were also identified in the service industry amongst the financial sector (Ashton, Maguire and Garland, 1982).

On top of the 'white heat' of technology that impacted the British economy since the 1970s three other significant trends are identified. The first relates to the relative decline in manufacturing and relative rise in service industries. From 1979, 7 million people were employed in the manufacturing industry, by 1983 that figure fell to 5.5 million. Concurrently, in 1979 employees in the service industry were at 13 million and rose to 15 million in 1989 (Ashton, Maguire and Spilsbury, 2016). The second relates to the growth of part-time employment (Furlong *et al.*, 2017). In 1971 15 per cent of all employees were working part-time, by 1989 this figure rose to 24 per cent (Ashton et al 2016). Thirdly, the increasing participation of women in the labour force from 38 per cent in 1971 to 45 per cent in 1987 (ibid). The growth of women in the workforce in connected to part-time labour - during 1986 women accounted for 80 per cent of additional part-time jobs (ibid). Whilst women have increased in terms of the overall share of employment, sex segregation is widespread due to the type of occupations and positions entered by women compared to men. The increased participation rate of women in the general labour market also presented a direct competition to youth labour (Ashton et al 2016; Furlong et al 2017).

The relative decline of apprenticeship schemes and increase in education opportunities due to the increasing pressure on young people to accumulate credentials resulted in a much higher proportion of school leavers in the 1970s onwards staying on within education than their earlier cohorts (Bynner and Ferri, 2003). Those who did not choose to stay on within education and had little to no qualifications faced the harsh reality of a ‘patchwork’ career trajectory, characterised by shifting occupations and periods of unemployment (Bynner, 2005). In 1976, the number of individuals who left school without qualifications was 21 per cent; in 1986, it was 9 per cent (Wallace and Cross, 1990). The 1970 cohort was the last to ever experience the dual O’level/CSE composition at 16 – the BCS cohort was in the middle of a massive amount of educational reform that would come in 1988 with the advent of the Education Reform Act. In particular, men saw a significantly increased probability of being in full-time education over employment compared to the 1958 cohort (Bynner and Ferri, 2003), though prominent members of men were also entering government training schemes like the YTS. For women, the decreasing numbers of young women being out of the labour force also saw a corresponding increase in labour market participation and higher education participation (Bynner and Ferri, 2003). The expansion of the university system in the late 1960s following the Robbins Report (Robbins Report, 1963) supplied higher education places that this new service-based labour market so often demanded (Bynner and Ferri, 2003). Compared to the continent at the time, European education participation rates were changing more rapidly than Britain (Bynner, Ferri and Shepherd, 2019). For most, the transition into adulthood is characterised by an initial movement from mandatory education to some form of employment. The fact that the BCS cohort appears to exhibit an elongated stay within education (Bynner et al., 2002) is some indication of the changing nature of the labour market within the UK – and also provides evidence for the development of an ‘Emerging Adulthood’ (Bynner, 2005). This transitional change is indicative of two potential sources; the first would be a significant economic shock in the form of a recession, which would encourage individuals to stay in education for longer to avoid the initial economic shocks and uncertainty that come with being employed in a labour market experiencing a downturn. The second relates to a degree of economic restructuring due to technological change, resulting in different skills and credentials, thus encouraging a more prolonged stay within education to garner such skills and credentials. The BCS cohort experienced two major economic shocks in their life course by age 16 – the 1973-5 recession and the 1980-1 recession. The BCS cohort also experienced the aftereffects of economic restructuring during the post-war consensus and a growing service economy (Bynner et al., 2002). Leaving school to enter employment for minimum school-age leavers was a much more complicated process compared to 10-20 years earlier – even more so for those living in industrial and manufacturing heartlands (Bynner et al., 2002).

The returns to education for the BCS cohort confer a 17 per cent average increase in income for those individuals who stayed on within education post-mandatory schooling compared to their peers (Boero et al., 2020). This is not entirely surprising, considering that education is the most important predictor of adult incomes and earnings (Breen, 2022). However, it does emphasise the importance of reflecting on the stratifying influences during education and their subsequent impacts on choice and opportunity post-education. This single most important predictor is a worrying phenomenon when combined with a ‘wastage of talent’ (Bukodi, Bourne and Betthäuser, 2017), whereby young people from disadvantaged backgrounds face barriers to fully realise their academic potential within the British educational system.

The changing role of education and individuals' relationship with it was not built-in isolation. The changing structure of the labour market also had other effects. Labour market restructuring was part of the increase in home ownership from the 1950s to the 1990s. In 1951, only 31 per cent of people owned their own homes; in 1991, this rose to 67 per cent (Bynner and Ferri, 2003). While homeownership increased, it was stratified by parental social class and income (Blanden and Machin, 2017). For the BCS cohort, having parents who were homeowners when they were aged 16 increases the probability of themselves being a homeowner at 42 by 116 per cent (ibid).

These historical phenomena have impacted the relative stability of youth transitions for young people in the BCS cohort. The relative decline in individuals moving straight from school into work after mandatory schooling and the growth of risky transitions and accumulating human capital via higher education suggests increased risk and uncertainty (Anders and Dorsett, 2017). The BCS cohort can be characterised as a group marked by dramatic and widespread societal change due to economic, political, and social upheaval. The changes and reforms as a consequence affected every aspect of the transitionary experience for young people.

#### Structural Barriers to successful transitions

##### Sex

Women's experience within the 1970 cohort saw a continuing weakening of gender differences in processes of occupational attainment – a similar trend seen within the 1958 cohort (Bukodi, 2009). However, the strength of education in this process appears to remain the same across cohorts (Bukodi and Goldthorpe, 2009). The weakening of gender differences is seen at the educational and occupational levels through take-home income (Bynner, 2005). However, whilst the BCS cohort experienced a decline in gender-segregated occupational sorting (Lekfuangfu and Lordan, 2022), occupations with the highest share of males maintained relatively high levels of segregation. Whilst it has been emphasised that social class origins have had an impact on the BCS youth, the changing nature of the labour market has also had ramifications for men and women concerning their biographical agency and their ability to find routes to stability and security (Schoon, Martin and Ross, 2007).

##### Social Class

The BCS cohort experienced a stratified post-mandatory schooling experience. Regarding participation in higher education, those from the most advantaged social origins were more likely to attend higher education institutions than those from less advantaged backgrounds (Alcott, 2013). Prior academic attainment explains most of the variance in this stratified higher education participation (around 60%) (Alcott, 2013). With the growth of an ‘Emerging Adulthood’ and an elongated stay within education, involvement in education for the BCS cohort has widened the gap between disadvantaged and privileged social origins (Bynner, 2005). These apparent returns to schooling are stratified according to social class origins, with the advantages offered by specific qualifications differing according to class origins (Bukodi and Goldthorpe, 2011; Parsons, Green and Wiggins, 2016).

##### Conclusion

The BCS cohort can be characterised by choice. Compared to previous generations, that choice was much more numerous in the options presented to the BCS youth on what to do after mandatory education. The ‘Emerging Adult’ could theoretically choose any of these options; however, the reality is that many of these options constrain the individual either immediately or down their life course. If the desired route from education were to find stable employment, the NCDS cohort would find that simply entering employment would provide a viable route to success. For the BCS cohort, however, this was not strictly the case. On top of a major recession, labour market restructuring and technological innovation provided a much more complex, elongated transition to a stable occupation (Martin, Schoon and Ross, 2008), resulting in a ‘winding road’ school-to-work transition (Leuze, 2010). Entering employment immediately after mandatory education could lead to periods of unemployment due to a lack of skills in a new economic landscape (Bynner, 2005). Joining a government training program like the YTS would provide some equally unsatisfactory results (ibid). Unemployment was a route that was even more restrictive than earlier cohorts due to cutting young people off benefits. Thus, the BCS cohort can be characterised as one of an educational turn. Staying within education, weathering the recession storm, and picking up relevant and sometimes required qualifications were most likely the best options to lead to a stable and successful occupational career. Unfortunately, education – particularly post-mandatory education – was highly stratified. This stratified nature impacted the most privileged – by giving them advantages in the labour market and the least privileged – by incurring further disadvantages. It should be assumed that, with this context, those individuals who entered education as a route post-mandatory schooling would thus be from more privileged backgrounds, perhaps in an even more striking ‘haves and have nots’ fashion than previously seen post-1944 Education Act reform.

### BHPS and UKHLS in Context

The BHPS and UKHLS unlike the other datasets mentioned thus far are not birth cohorts, as such individuals taken from these datasets are not all born within the same year. The BHPS and UKHLS synthetic cohorts conducted their first youth transitions within the 1990s and 2000s. The labour market continued its trend of declining manufacturing and the education system continued a reformation process along the lines of the Education Act in 1988 to further widen particiaoption following an established credentialled labour market. The youth of this timeframe sat examinations at age 16 – unlike the O’levels of the NCDS and BCS cohorts, members of the BHPS and UKHLS synthetic cohorts were expected to sit GCSEs. Youth training underwent constant reforms, replacing the YTS with Youth Training and eventually the New Deal for Young People. Whilst change is evident across these synthetic cohorts, the purpose of youth training schemes in these cohorts follows a similar approach to those of the BCS timeframe. The developments and changes to British society can be characterised as a continuation of change rather than a drastic deviation from it of earlier cohorts.

#### Story of BHPS and UKHLS youth

Youth training continued to be altered and reformed in the 1990s with the YTS being replaced by Youth Training (YT) in 1990 (Droy, Goodwin and O’Connor, 2019). Following the consistent levelled critiques against the YTS scheme, it was eventually reformed into the Youth Training (YT) scheme in 1990 following a process of devolution of youth labour intervention, following the establishment of the Training and Enterprise Councils (TECs) in 1982 and the National Training Task Force (NTTF). The YT scheme was essentially the same as the YTS scheme with greater flexibility on types of schemes offered, and the time period of each scheme provided (Deakin 1996). The YT, like the YTS was heavily critiqued, particularly for its failure to deliver on its main purpose - training and finding placements for young people. A survey conducted in 1995 demonstrated that up to 60 per cent of people that participated in the programme had no qualifications upon leaving (Furlong et al 2017).

The Major government established the 'Modern Apprenticeship' in a bid to provide better quality training (Furlong et al 2017). This, like other programmes described, was a failure. Whilst attempting to model the modern apprenticeship scheme closer to the German model, critiques argued that it was more like a rebranded YT scheme (Ryan and Unwin, 2001). Only one in two school leavers on the modern apprenticeship scheme attained a NVQ level 3 qualification - the purpose of the programme (Furlong et al 2017). Further critique came from the strict gender segregation of the programme, Women were placed in health and social care as well as customer service at a rate of 86 and 69 per cent compared to engineering and construction roles at 3 and 2 per cent (Furlong et al 2017).

The rise of the Blair administration inf 1997 came with it another attempt at providing adequate training schemes to young people. Launched in 1998 the New Deal for Young People (NDYP) provided unemployed youth a six-month training course with tailored support using a personal advisor. A variety of options were provided to the young person, they could: be given a subsidised job placement for six months, a place on full-time education courses up to 12 months, or a work placement with the environmental task for or community task force for six months. The NDYP was certainly an improvement on past schemes, though its effectiveness has been left to debate, considering youth unemployment levels were steadily decreasing from the introduction of the scheme (Furlong et al 2017).

The Coalition government in 2011 launched their newest flagship scheme called the Work Programme. Unlike others before it, the programme is contracted completely out to private and third sector organisations (Furlong et al 2017). The Work Programme has been an abject failure - of the 785,000 young people in its first year, only 2.3 per cent held a job for six months or more (Murray, 2012). The programme continues a tradition started in the Thatcher administration of withholding benefits if the individual refuses to participate. It has been argued that more people will have been sanctioned by the Work Programme than properly employed through it (Furlong et al 2017).

Very few young people from the 1980s onwards left education without completing upper Secondary Education (ibid). In the UK between 1984 and 2013 full-time participation among 16–24-year-olds increased from 1.42 million to 3.03 million. The expansion of educational participation was a direct response to a lack of labour market opportunities for young people immediately following the end of mandatory education.

Due to this widening participation, youth that experienced the 2008 recession were much more highly educated than their peers facing the 1970s and 1980s recessions (Bell and Blanchflower, 2013). This increased level of educational qualifications has done little to change the direction or concentration of labour market positioning for young people. Young people are still heavily concentrated in low skills service labour areas of the economy, more than seven in ten 16–24-year-olds work in elementary occupations (Office for National Statistics, 2014). The continued rise of the service sector of the economy - 21.5 per cent from 1979 to 2010 - has resulted in youth labour being restricted to low skill, low pay occupations where in the two years between 2010 and 2012, 77 per cent of new jobs were in low paid occupations (Spence, 2011; Trades Union Congress, 2013).

## Breakdown of Dataset Context

This section details the description of each cohort’s data and structure. This section provides a description of each dataset, its construction, and how analysis will follow going forward. Following from this, each variable will be introduced and provided background information on for each dataset. This section ends by a detailed description of the analytical models that will be used in this section.

### NCDS Dataset context and methodology

The National Childhood Development Study is a birth cohort survey with a survey design that is relatively straightforward to derive an analytical sample. The NCDS does not have any weights, nor does it oversample. The construction of a given analytical sample using the NCDS requires the merging of the relevant waves of data using the unique identified [ncdsid] to link observations to individual strata. For this analysis data is merged from waves 0-4, or from ages birth to 23 to construct an analytical sample that includes data on youths first major transition post-mandatory schooling that includes measures of sex, housing tenure, parental social class. The total complete records cases for the proposed NCDS model are N=12,411.

### BCS Dataset context and methodology

The British Cohort Study is like the NCDS a birth cohort survey with a survey design that is relatively straightforward to derive an analytical sample. Like the NCDS, the BCS does not have any weights, nor does it oversample. The construction of a given analytical sample using the BCS requires the merging of the relevant waves of data using the unique identified [bcsid] to link observations to individual strata. For this analysis data is merged from ages birth to 30 to construct an analytical sample that includes data on youths first major transition post-mandatory schooling that includes measures of sex, housing tenure, parental social class. The total complete records cases for the proposed NCDS model are N=1,574.

There are substantive differences between the BCS dataset and NCDS dataset. This difference relates to the construction of social class measures (NS-SEC and RGSC). Whilst both the NCDS and BCS datasets use occupational coding data from (Gregg 2012), the NCDS codes are only available for fathers of participants, while for the BCS cohort, both fathers and mothers are made available. Due to this, both NS-SEC and RGSC are coded by using mothers’ occupational data to fill in any missing data entries from the father’s data. Besides these two differences, the model presented for analysis in chapter two is identical to that of chapter one. This is to start to build a historical picture of the changes and developments in choice and opportunities for different cohorts across different periods.

### UKHLS Dataset context and methodology

Unlike the birth cohort studies the BHPS and UKHLS datasets have a different survey design. Rather than having all data within one ‘master’ wave dataset like the cohort studies, the BHPS and UKHLS breakdown their individual waves by key topics, focusing on household or individual level effects. The main datasets that will be used in this analysis come from the youth and indresp datasets. Each wave file is denoted with a wave suffice “x\_”. The youth panel collected all data on individuals within the BHPS and UKHLS that were between the years of 11-15 years old. The youth panel started at wave 4 of the BHPS or in 1994. Interviews were conducted face-to-face. The inderesp panel collected data on the entire adult population of the BHPS and UKHLS surveys. The youth panel did not directly collect any data on variables of interest to the proposed analysis. It’s primary use in this analysis comes from its ability to construct synthetic cohorts using birth date data from individuals within the youth panel. Isolating young people into school leaving years for each subsequent wave of the youth panel allows the construction of synthetic cohort data. Using individuals unique cross-wave identifier and cross-household identifier, the mothers and fathers of youth can be identified, and their data also used within the analysis. Synthetic cohorts are constructed at time *t* and merged with indresp datasets at time *t+1.* The youth panels maximum age is 15, there is no adult level data on individuals at age 15 so to get access to said individual level data a wave needs to pass so they enter into the adult survey. For example, young people in the youth panel at age 15 in wave 4 of the BHPS will be merged into the inderesp dataset at wave 5 of the BHPS to collect adult level individual data such as current economic activity.

As mentioned previously, through this merging process the mother and father personal identifier within the household can also be identified. From this, parental variables can be constructed. As well as household level effects like housing tenure.

There is a possibility that when merging the youth panel into the indresp data that missing data may occur due to survey construction. Some variables in the BHPS and UKHLS are not asked at every survey wave. An example would be a variable that is only asked at wave 1 and wave 12 of the UKHLS. For young people entering the UKHLS at a later stage than wave one this information is not immediately available. With fixed intrinsic characteristics like race or sex, using the multiple waves of the household panel survey, the personal identifier pidp can be used to find any recorded instance of a value within the dataset and change that value in the current instance from missing to the corresponding matching value at a later or earlier date. If for example a young person enters at wave 4 of the UKHLS there would be no value for them, however, when it is asked again in wave 12, there should be a value. By replacing the missing value in the past with the current value at wave 12 it alleviates the amount of missingness within the given sample.

The result of this forms the BHPS and UKHLS synthetic cohort sample. This process is documented in figure 2.1.

Table 2.1 Synthetic Cohorts using BHPS and UKHLS samples

A black background with white ovals

Description automatically generated

### Introduction to Measures for Subsequent Analysis

The following section provides an overview of key variables used for this analysis. For this analysis, sweeps 0-4 (up to age 23) will be used for the NCDS, sweeps 0-5 (up to age 30) will be used for the BCS, and sweeps from BHPS wave 4 up to wave 13 of the UKHLS will be used.

From this variable selection measures related to economic activity, educational attainment, sex, social class, and housing tenure were selected for inclusion in subsequent analysis. This section will follow sequentially a discussion of the given raw variables for construction and re-coding for each dataset, starting with the NCDS, then the BCS, and finally the UKHLS.

#### Economic Activity

The primary outcome variable of interest for cohorts is the main economic activity of individuals after mandatory schooling. For the birth cohorts the September when they are 16 is selected -- month 201 since birth – this translates to the month of September when all cohort members are aged 16. For the synthetic cohorts there is no monthly economic activity recorded in the household panel surveys. Instead, there is a single economic activity question asked for all individuals at each wave of the BHPS and continued into the UKHLS.

The economic activity variable records what cohort members were doing after they had left mandatory schooling in September at age 16. For example, the economic activity individuals engaged with after year 11 in the English and Welsh school system context. September was selected to allow time for youth to gain their examination results.

The economic activity variable [ec201 (for NCDS)], [JACTIV + va86sep (for BCS)], [x\_jbstat] (for UKHLS)] was collected at different times and different ways for each of the datasets used for analysis. The subsequent section will detail, starting with the NCDS how each variable was collected, and how it is used.

The NCDS economic activity variable [ec201] was a retrospective work history collected at age 23. Participants were asked to note their current economic activity from age 16-23 each month. This variable comes from sweep 4 (Age 23) of the NCDS. The analytical sample’s economic activity was recorded retrospectively by the participants at age 23 each month from when they turned 16 to when they turned 23. Information for the following variable comes from the data dictionary part 1 (National Children’s Bureau, 1981). Each month is recorded as a diary that covers one possible main activity defined as ‘Jobs’, ‘Full-time Education’, ‘Unemployment’, ‘Out of the labour force’, and ‘Fill-in-time’[[10]](#footnote-10). The monthly diary of economic activity filled out by participants was coded by a coder.

Table 2.2 Frequency Statistics for Economic Activity

|  |  |
| --- | --- |
|  | Frequency |
| Economic Activity in Month 201[[11]](#footnote-11) |  |
| Missing | 86 |
| Full Time Job | 4,716 |
| Full time job + part time education | 144 |
| Full time job + full time job + apprenticeship | 1,842 |
| Full time job + apprenticeship + part time education | 22 |
| Full time job + apprenticeship + day block release training course | 21 |
| Full time job + other training course | 1 |
| Full time job + day block release | 366 |
| Full time job + day block release training course + part time education | 4 |
| Full time job + other | 20 |
| Full time job + other training course + part time education | 1 |
| Full time job + full time night training opportunities for young parents training course | 35 |
| Full time job + full time training course + part time education | 1 |
| Full time job + local government support scheme | 2 |
| Full time job + local government support scheme + day block release training course | 1 |
| Part time job | 37 |
| Part time job + part time education | 2 |
| Part time job + day block release training course | 2 |
| Training opportunities for young parents | 1 |
| Local government support scheme | 1 |
| Full time post school education | 1,046 |
| At School | 3,717 |
| Unemployed | 276 |
| Unemployed + part time education | 3 |
| Unemployment Rule 6 | 11 |
| Out of the labour force | 164 |
| Out of the labour force + part time education | 3 |
| Part time education | 11 |
| Total | 12,536 |

The original economic activity variable for month 201 has 28 unique values. These 28 values comprised a combination of main categories: employment, education, training, and unemployment. Individuals could, for example, be coded as being in full-time employment and doing an apprenticeship scheme, etc. These 28 unique values were recoded into a simple binary variable of ‘continuing schooling’ and ‘not continuing schooling’ to provide the most straightforward analysis of an individual’s first transition.

Moving on to the BCS construction of economic activity, individuals’ activity was recorded retrospectively in the 21-year-old sub-sweep as well as the economic activity dataset whereby retrospective data was collected for work and non-work activities since the age of 16 at age 30 sweep - this will be discussed shortly. The 21-year-old sub-sweep contained 10% of the participants in the sample. The 21-sub-sample sweep was drawn from cohort members who are residents in England and Wales – no data on Scottish residents was included; interpretations of data using this sweep are restricted to England and Wales only. Interviews were conducted face-to-face in 25 clusters based on 26 postcode areas (Bynner, 2017).

The original raw economic activity variable [va86sep] in the 21 sub-sample sweeps is provided below in Table 2.4 13 cases were dropped as missing due to being categorised as ‘Something Else’. Documentation did not provide any detailed insights as to what this category meant and as such observations could not be assigned to any category.

Table 2.3 Frequency Statistics for Economic Activity

|  |  |
| --- | --- |
| Economic Activity in Month 201 | Frequency |
| Seeking Work | 68 |
| Looking After Children/home ft | 9 |
| On Training Scheme | 426 |
| FT Education | 723 |
| FT Employee | 352 |
| PT Employee | 35 |
| Self-employed | 5 |
| Something Else | 13 |
| Total | 1,631 |

Another retrospective economic history was taken within a special ‘Economic Activity Dataset’. This attempted to take retrospective diaries on all observations within the sample from 1986 – 2013. This is the only other economic activity data that is provided by the BCS. The original raw economic activity variable [JACTIV] is provided below in Table 2.5. There are some issues with the JACTIV variable and more widely the economic activity dataset. In some instances, individuals are coded as having multiple main economic activity statuses within the same time instance. For example, out of the 2,374 individuals that have economic activity status for the month of September following the end of mandatory schooling at age 16, 69 of these observations are repeated more than once – often times with different economic statuses. The majority of these overlapping activities correspond to individuals undertaking full-time employment as well as full-time education. To solve for this issue, a secondary variable within the dataset [JDUR] measures the length of time in months an individual is engaged in any given economic activity spell. Of the 69 overlapping cases, the longest spell in months was taken to be the dominant spell and used as that observation’s ‘main’ economic activity.

Table 2.4 Frequency Statistics for Economic Activity Part 2

|  |  |
| --- | --- |
| Economic Activity in Month 201 | Frequency |
| Don’t Know | 8 |
| FT Employee | 876 |
| PT Employee | 35 |
| FT Self-Employed | 25 |
| Employed | 7 |
| Unemployed | 34 |
| FT Education | 1005 |
| PT Education | 8 |
| Government Training Scheme | 355 |
| Permanently Sick | 3 |
| Looking After Family | 7 |
| Retired | 2 |
| Travelling | 1 |
| Other | 8 |
| Total | 2,374 |

Both constructions of economic activity are required for the analysis of the BCS cohort due to the low number of observations recorded. Both variables have an overlap of 320 cases. Using the variable constructed at the 21-year-old sub-sample as a base variable, the variable constructed using the economic activity history is used to fill in missing data if and only if data is not already available from the 21-year-old sub-sample. The former was used as the ‘master’ variable because the data was collected closer to the time of the economic activity occurring compared to the other variable and as such there is a greater likelihood of more accurate retrospective collection of economic activity data at age 16. After this merging, the economic activity variable was re-coded along similar lines within the NCDS model, into a simple binary variable of ‘Continuing Schooling’ and ‘Not Continuing Schooling’.

The BHPS and UKHLS surveys do not collect economic activity data on the youth panel (individuals below the age of 16). To ascertain youth economic activity data the youth panel is required to be linked with the indresp files within the BHPS and UKHLS datasets for each wave. For a young person that is 15 in the fourth wave of the BHPS, the fifth wave of the indresp files will be used to match the id [pidp] of an individual that is now 16 and their economic activity data is now collected.

For the BHPS the survey data was collected between September-December of each survey year – this makes it an adequate direct comparison measure for the birth cohorts. The UKHLS has a more complex survey design that collects survey data for a given wave over multiple years. Individuals asked about their economic activity status in the UKHLS could be a minimum of age 16 in January of that survey wave and a maximum of 26-30 months older dependent upon the particular survey wave. Observations are thus restricted for 16-year-olds during or after the month of September to avoid counting cases whereby youth are still in mandatory schooling but are the oldest in their school cohort year – this avoids counting cases that would 100 per cent count as ‘schooling’.

The BHPS and UKHLS surveys provide a variety of categories that measure economic activity. Restricting observations to the youth panel only presents a table of descriptive statistics for the raw dependent variable in table XXXX. [Need UKHLS data before completing section].

#### Educational Attainment

The NCDS cohort members reached the compulsory school leaving age in 1981. At this time, the primary educational qualifications were either the Certificate of Secondary Education (CSE) (Pearson qualifications, 2023a), introduced in 1965, or the Ordinary level or O’level, introduced in 1951 (Pearson qualifications, 2023b). The O’level was understood to be a higher level than CSEs, and fewer people achieved O’level grades. This is the best and most advanced ability measure for the age of 16 and makes a good measure of educational attainment for those at 16 after mandatory schooling ends.

The educational attainment variable is constructed in a binary less than five O’levels/five or more O’levels variable. Within contemporary literature on educational attainment, gaining five or more GCSEs at grades A\*-C is a standard benchmark measure used within official reporting (Connelly, Gayle and Paul S. Lambert, 2016).

There is an argument that GCSEs and O’levels are analytically distinct concepts, and as such, a like-for-like measure may not be the most attractive (Murray, 2011). A measure of attainment, GCSEs and O’levels provide considerable barriers to entry for young people pursuing future goals (ibid). Due to this rationale, using a threshold measure for number of O’levels given the restriction of age on the amount of attainment an individual could have undertaken at this time, it appears to be the best operationalisation of the measure. For this reason, it is rationalised to prefer the five or more measures used within GCSE-based literature for O’level attainment.

This variable was constructed from two separate variables – the first was a simple binary variable of whether an individual had any O’levels [n4655], the second, on condition of the first, then asks how many O’levels that person had passed [n4656] – passing in this context refers to if an individual’s O’level grade was within the grade boundaries A-C. For context both variables include Scottish O’grades within the O’level variable construction. Combining these two variables produces a single count variable that includes the number of zeros. This attainment variable was then recoded into a binary variable of less than five O’levels and greater than five O’levels. This was done for two reasons. The first has been discussed above. The second reason for recoding is one of practicality. Keeping O’levels as a count variable illustrates a truncated position of several O’levels, making a binary dummy more sensible.

Moving on to the BCS construction of educational attainment, the BCS cohort members reached compulsory school leaving age in 1986. The BCS cohort was the last group to experience the O-level/CSE split system (Pearson qualifications, 2023b). The variable itself is a binary variable of the number of O-level passes. The construction of educational attainment in the BCS cohort is complicated because attainment for individuals was first coded when participants were 26 years old and then once again at age 30. At that point in the cohort, only 9,003 participants responded, of those, only 5,438 responded to an educational attainment variable.

The BCS at age 26 documents O’level attainment in two ways. The first is a variable of the number of O’level passes ranging from A-C grade [b960154]. The second is a variable of the number of O’level passes going from D onwards [b960157].

The educational attainment variable takes all data from the variables where individuals received A-C grades and codes that into a binary less than five/five or more variables. There are instances where data is missing in the former variable but available in the latter. In these instances, it is assumed that individuals only received ‘other’ O’level grades. When this is the case, this data is coded as individuals receiving less than five O'levels. Unlike the NCDS, where O’level passes were coded for all individuals in the UK (Scottish equivalents were automatically coded into the O’level variable), the BCS data separates Scottish educational data from the rest of the UK. This meant that the Scottish equivalent for O’levels at the time of the BCS – Ordinary Grades, or O’grades, were merged with the original O’level passes variable. The procedure for dealing with Scottish grades was identical to that for O’level grades. It had two variables: one that hosted the number of O’grades A-C [b960169] and another that hosted the number of O’grades D-onwards [b960172]. All these variables were combined to make an O’level attainment variable as a measure of educational attainment. However, this only accounts for 5,438 individuals in the total cohort. At age 30, individuals were again asked to record their educational attainment and number of O’level passes. This is merged with the educational attainment variable to boost observations. The educational attainment variable takes a semi-dominant approach to this merging. The underlying thought process is that at age 26, an individual will be more likely to accurately recall their educational attainment than at age 30. Thus, in cases where there are repeated observations and they differ, age 26 is given dominance. At age 30, a variable [edolev1] gives a count of the number of O’level passes. Unfortunately, at age 30, the BCS decided not to document how many O’grade passes Scottish students attained – instead opting for a simple ‘Did you complete a Scottish qualification’ variable. This could lead to a substantive amount of missingness amongst Scottish individuals.

Finally, moving on to the household panel surveys. The BHPS uses two count variables referring to the number of GCSEs grade A-C [nqfede] and the number of GCSEs grade D-G [nqfedd]. These two variables were asked as every wave of the BHPS starting with wave 1. For the youth panel merged into *t+1* indresp files and interview dates on or past September of their 16th birthday the variables are used like in prior constructions, to create a simple dummy variable of: five or more GCSEs A-C versus less than five GCSEs A-C. Where there are instances of individuals have observations on the latter [nqfedd] variable but not the former [nqfede] variable, that is taken to assume that individuals only received grades at D-G grade and are thus coded as receiving less than five GCSEs A-C. The UKHLS sample does not collect any educational attainment related data. This is because the UKHLS was linked with the National Pupil Database, an administrative linkage that hosts all academic qualifications and information’s regarding UKHLS sample members. Using secure data access educational qualifications were isolated and coded in the same binary format as the BHPS sample, the two were then combined to form a total sample variable. [MORE NEEDED HERE AFTER SECURE ACCESS DATA].

#### Sex

Sex is a variable derived from sweep 0 in the NCDS [n622\_4]. Its inclusion for analysis is because, during the timeframe of the cohorts studied, sexed dynamics played an essential role in economic activity (Dex, Ward and Joshi, 2008; Bukodi and Dex, 2010; Dex and Bukodi, 2012). The evolution of part-time work, the differences in populations out of the labour force, and those choosing to go on to higher education are all influenced by sex (see Jones, 1986). For these reasons, sex provides a theoretically compelling case for inclusion within a model of economic activity post-mandatory schooling.

The BCS construction of sex was slightly complicated by the fact the BCS cohort boosted additional sample members beyond the birth sweep. This mean that a simple extraction of a birth sweep sex variable was not possible and had to be supplemented with a sex variable from other sweeps. Sex as a variable is taken at birth [a0255], though not all people included in the following sweeps have data for sex available; thus, this original sex variable is supplemented through a variable at age 26 [b960337] and age 30 [dmsex].

As for the BHPS and UKHLS, all individuals within both datasets have sex recorded in a master ‘xwavedat’ file. Using individuals’ unique personal identifiers [pidp] a sex variable was constructed for all individuals with no missing data using the [sex] variable.

#### Race

Echoing the arguments for the inclusion of sex in models of analysis, the role of race also impacted individuals’ economic activity during the NCDS timeframe (Payne, 1995; Lindley, 1996; Gayle et al., 2009). Race as a variable [n2017] for inclusion in this model presents too many statistical issues to be an effective measure.

This is because in the NCDS white people make up 96% of all participants. The resulting ethnic minority categories are thus too small to conduct helpful analysis. Initially, the resultant variable was parametrised as ‘white’ and ‘non-white’. Two significant issues resulted in the race variable being dropped from the analysis. The first returns to the overall low sample of non-white participants when spread over five different economic activity sub-categories. This low number of observations results in low statistical power and thus would impact the entire model. The second is that missing data is a particular problem regarding race. The race variable accounted for 16 per cent of missingness in subsequent models. On top of these two primary concerns, a combined race category into white/non-white presents assumptions surrounding homogeneity within the non-white category that is not theoretically justifiable (Connelly, Gayle and Paul S Lambert, 2016). For the same reasoning the BCS also does not offer a practical inclusion of a race variable for analysis. In keeping with the desire to provide models as close to identical as possible, race also was not included in the orignal pooled models using UKHLS data either. Within the detailed analysis of the UKHLS data however race will be used to compare model fit and the impact of race. The BHPS sample collects data on race at every wave in the inderesp files [race]. Race is also recorded in the xwavedat files [race\_bh] for the BHPS. For the UKHLS race is recorded in indresp, youth, and xwavedat files [ethn\_dv].

#### Housing Tenure

Previous analyses have used housing tenure regarding educational attainment and labour market outcomes (Di Salvo and Ermisch, 1997; Duta et al., 2021). Housing tenure is the primary structural inequality pointed to by proponents of new structuralism that is thought to have eclipsed traditional inequalities such as social class and (Devine, 2017). The argument proposes that because of an increase in social mobility, decline in trade union membership and a growth of cross-social class-based families this had undermined the power of social class (Devine, 2017). In its place, other structural cleavages such as housing tenure are on the rise in terms of how much of an impact it has upon individual choice and opportunity. Housing tenure enables the inclusion of a ‘consumption cleavage’ (Saunders, 2003, 2021) based variable. This form of cleavage encapsulates the ‘new structuralist’ notions that structural inequalities do, in fact, matter, but not necessarily older structures such as class and gender. Including housing tenure in this model allows a more direct investigation of this sentiment. For subsequent analysis, tenure measures whether an individual lives in a home owned by their parents or not within the NCDS [n1152][[12]](#footnote-12).

For the BCS, housing tenure was taken from when respondents were ten years old. There was information on tenure when respondents were 16, although the responses were scattered across several binary variables with low overall responses. The age ten variable on housing tenure was a multiple-category variable with few overall missing cases. Housing tenure as a measure for inclusion in this model is the most critical measure to focus upon. The arguments of Saunders (2002, 2003, 2021) and other new structuralists were born when the BCS cohort was economically active. Statements related to the ‘death of class’ (Pakulski and Waters, 1996) and the rise of tenure as the most substantive structural explanation for inequality are central to this chapter. As a variable housing tenure is taken at age 10 [d2] – like the NCDS cohort, this is again like other variables supplemented by a set of variables on housing tenure at age 16 [of3\_1, of3\_2, of3\_3, of3\_4, of3\_5].

In 1986, home ownership rates within England stood at 63.5 per cent (HomeOwners Alliance, 2012). Within the BCS, the recoded housing tenure variable has 61.87 per cent of the sample owning their own home compared to 38.13 per cent that do not own their own home. This is relatively like the official statistics of England at the time in 1986. However, when fitting housing tenure into the complete records analysis for model interpretation, missingness from other variables causes a shift in these per cent rates. Within the Complete Records model, those that own their own home jumps to 78.98 per cent compared to 21.02 per cent that do not own their own home. This is over 16 per cent larger than the official statistics report. This is an initial indication that missingness may pose a problem for the model in the future and that specific techniques, such as multiple imputation, may be helpful to investigate this.

Finally with respect to the BHPS and UKHLS both use the same housing tenure variable collected across all waves of each survey located in the hhresp and indresp dataset files [tenure\_dv]. This is a derived variable that uses variables [HSOWND, HSJB, RENTLL, RENTF]. These variables relate to if an individual owns their own home, accommodation with present job, landlord of rented accommodation, and furnishings in accommodation. For this household level variable instead of using the young person’s pidp, instead the parental pidps of mother and father will be used to provide a value for youths housing tenure.

#### Social Stratification Measures

Social stratification is a cornerstone of sociological research. No one universally agreed measure of social stratification has been selected as the measure of social stratification. There are many schools of thought when attempting to capture social stratification – for the sake of this thesis, two will be focused upon in detail. The first is a measure of social class, which contemporarily employs an occupation-based schema. The second is social stratification scales, which instead rely on capturing a continuous measure.

Social class as a variable has constant and consistent debate throughout sociological literature (Bottero, 2004) – even today, whilst current schemas reign dominant, there is no agreed upon universal measure of social class. Three social stratification measures of NS-SEC, RGSC, and CAMSIS will be used due to their theoretical distinctiveness and empirical operationalisation.

A vital aspect of this chapter is comparing the substantive findings of models with different social stratification measures to see if there are different patterns for different dimensions of social stratification. Multiple measures of social stratification are reflected upon. The following section seeks to establish the significant measures of social class and weigh their common strengths and weaknesses, which may affect model parsimony. All models will be compared and interpreted, with the intention to understand how different measures of social stratification may impact the substantive interpretation of a given model. Goodness-of-fit statistics will be discussed and the ‘best’ model with the lowest AIC and BIC will be selected for further inspection unless there is adequate justification to select another model instead.

Longer-term structural transformations of society will alter the underlying distribution of stratification over time (Lambert and Barnett, 2021). Whilst the Treiman constant - a concept that argues that occupational positions have the same meaning over time and across different countries meaning that hierarchies of occupations are for the most part invariant across multiple different societies and across time (Treiman, 1977) - is often hailed as the single most crucial empirical generalisation to be confirmed through social stratification research (Lambert *et al.*, 2008) and thus justifies the using universal and semi-universal social stratification coding of occupational data. Structural transformations over time (ibid) can potentially alter the underlying distribution within these universal and semi-universal coding schemas.

The NCDS provides occupational coding measures for the father’s socio-economic position using a variety of measures (Gregg, 2012). The measures provided are the Registrar General Class Schema (RGSC), National Statistics Socio-Economic Classification (NS-SEC), and the Cambridge Social Interaction and Stratification Scale (CAMSIS). Occupational codes were constructed for fathers of NCDS youth in 1969. This was based upon text-based responses from the parental questionnaire concerning parental employment and occupation. Unfortunately, no such occupational measures were taken for mothers, making it impossible to employ a semi-dominance approach (Connelly, Gayle and Paul S. Lambert, 2016). The occupational coding conducted by (Gregg, 2012) created occupational code classifications using response text strings using the Computer Assisted Structured Coding Tool (CASCOT), following this SPSS syntax was used to automatically convert the CASCOT codes into social class based on the RGSC and NS-SEC. CAMSIS codes were produced the same way. This thesis has conducted its analysis using Stata, the conversion of the SPSS and Excel files from Gregg appears to have created errors with the produced data – particularly the CAMSIS codes. Where CAMSIS is supposed to have a mean of around 50 and a standard deviation of 15 the codes produced by Gregg have a mean of 4.44 and a standard deviation of 10. It 4appears in the conversation the CAMSIS codes have had their decimal point shifted one place to the left. Because of this error I manually re-constructed all social stratification measures using the original Gregg SOC codes and then compared all three measures to the ones constructed by Gregg. For CAMSIS there was an obvious difference, but RGSC and NS-SEC also displayed minor differences also. Going forward with the analysis the manual re-constructions were selected over Greggs produced code.

The variables used to construct the social stratification measures for the NCDS are for NS-SEC [N2SNSSEC], for CAMSIS [N2SCMSIS], and for RGSC [N2SRGSC]. Manual re-construction of these variables used standard occupational codes provided at SOC 2000 [N2SSOCC] and SOC 90 [N2SSOC90].

The BCS uses the same occupational coding conducted by Gregg (ibid) as the NCDS. As such manual re-coding was also required for construction of social stratification variables for the BCS also. The BCS construction of social stratification measures is slightly different from the NCDS construction due to the provision of new data. While the basis of each measure is the father’s social class position when the respondent was ten, missing responses are filled in with the mother’s social class position when the respondent was ten. This accomplishes three things. The first is that the mother’s social class position fills potential item missingness. The second is that it offers those respondents who do not come from a traditional nuclear family the ability to enter the model by taking the mother’s social class position where a father’s is not present. Finally, through both accomplishments, the level of missingness and overall responses is increased within the model, enhancing the statistical power of the model overall.

All social stratification codes are taken from Gregg’s documentation of the NCDS and BCS (Gregg, 2012). For the NS-SEC construction, a semi-dominance approach was used. Thus, a variable on the father’s NS-SEC position at age 14 [B3FSNSSEC] and the mother’s NS-SEC position at age 14 [B3MSNSSEC] are combined (when the father’s data is not available, the mother’s data is used instead. The same procedure is used for the RGSC construction, with the father’s data [B3FSRGSC] and the mother’s data [B3MSRGSC]. The same procedure was used for CAMSIS, the father’s CAMSIS at age 14 [B3FSSOCC] and the mother’s data [B3MSSOCC]. Similarly to the NCDS construction of CAMSIS, the files produced by Gregg (Gregg, 2012) erroneously erased the qualitative distinctions in CAMSIS by shifting the decimal point one to the left and rounding up to one decimal point. CAMSIS was reconstructed using standard occupational codes provided by Gregg (ibid) at SOC 2000 [B3FSSOCC] and SOC 90 [B3FSSOC90] for the father and at SOC 2000 [B3MSSOCC] and SOC 90 [B3MSSOC90] for the mother.

The overall patterns of social class position between the NCDS and BCS have not changed substantively. This lack of change is worth noting, considering the relatively large-scale changes the British economy and society underwent during 1958-1980 (Bynner *et al.*, 2002; Furlong and Cartmel, 2007; Jones, 2016). Below is a table comparing each social stratification measure between the NCDS and BCS cohorts.

Both the BHPS and UKHLS require at least special licence access to use designated four-digit SOC codes using variable [jbsoc00] for SOC 2000 codes. The BHPS also offers SOC 90 codes [jbsoc90] and the UKHLS sample provides SOC 2010 codes in the form of [jbsoc10]. The BHPS sample does not have SOC 2010 codes and the UKHLS sample does not have SOC 90 codes. Unlike the prior birth cohort studies a detailed analysis of the UKHLS will not involve a sensitivity analysis of SOC measures. Unlike the NCDS and BCS cohorts the BHPS and UKHLS samples also have access to the data necessary to construct full versions of NS-SEC, CAMSIS, and RGSC. Employment status – used to derive the full socials stratification variables schemes uses the same variables across BHPS and UKHLS waves and is asked every wave starting with the first BHPS wave. Employment status uses [jbsemp, jbmngr, jbsize, jssize, jsboss]. Combined with SOC 2000 [jbsoc00] and all three social stratification variables can be derived. It is worth noting that both the UKHLS and BHPS samples have a pre-existing derived current NS-SEC job variable [jbnssec\_dv] however they do not offer a CAMSIS or RGSC measure. For this reason, all measures will be constructed using the same variables following correct construction procedures. An analysis comparing a simple construction versus full construction will be found in the detailed analysis section of the UKHLS instead of a comparison of SOC constructions like NCDS and BCS samples.

The following variables used for subsequent analysis are all considered variables of social class or social stratification scales and all use Standard Occupational Codes (SOC) as part of their construction. Two are social class schemas, and one is a stratification scale. By comparing cohort substantive findings, multiple socioeconomic measures are considered to see whether there are different patterns for different dimensions of social stratification both within cohorts and across them.

The following section provides a detailed breakdown of each chosen social stratification measure, explaining its theoretical makeup and analytical construction.

##### SOC Codes

Whilst all three stratification measures have different theoretical underpinnings, all three are occupational based measures. All three measures rely in part upon occupations to ascertain an individual’s position. All three use the same Standard Occupational Classification System (SOC) codes. SOC codes have gone through four different manifestations – starting in 1990 with the first SOC 90, and then being edited every ten years subsequently to keep the codes up to date and in line with contemporary British society. The occupational files produced by Gregg (2012) provide both SOC 90 and SOC 2000 codes for the Fathers of the NCDS cohort. In isolation there is a strong argument to use SOC 90 codes to produce the various social stratification measures for the proposed model. SOC 90 codes are closer in time to the NCDS and BCS compared to SOC 2000. This closeness in time should also present a more accurate portrayal of British society in comparison to the SOC 2000 codes. The reverse being true for the UKHLS and SOC 2000 codes.

The duplication of analysis using subsequent databases – the BCS and UKHLS – presents an issue. Whilst for example SOC 90 codes may be better suited for an analysis of the NCDS, the opposite is the case for a database such as the UKHLS in its later waves. Using different SOC codes for subsequent models would change the composition of the social stratification measures and is an inadequate response. Therefore, an argument for using SOC 90 codes for all models or SOC 2000 models presents itself. Whilst the former argument favours the older databases such as the NCDS and BCS the latter argument favours newer databases. Both arguments are fundamentally one of harmonisation. The proposed solution is to produce a sensitivity analysis on each social stratification measure using both SOC 90 codes and SOC 2000 codes and then use goodness-of-fit statistics to determine the most parsimonious model for each dataset whilst providing all alternatives. This solution also provides an ability to garner insight into the level of substantive difference that may or may not occur from using different SOC codes to construct different measures of social stratification.

The reason for SOC 2000 was predicated on the supposed need to improve alignment with the International Standard Classification of Occupations (ISCO) and the need to restructure the occupational classification based on the need to create NS-SEC (*SOC 2000 - Office for National Statistics*, 2000). Whilst the former does not matter for research such as this that does not use cross-country comparisons, the latter proposes another interesting reason to conduct a sensitivity analysis of SOC codes. The fact that SOC 2000 was produced not only to update occupational classifications to be more in line with the reality of contemporary Britain, but also to aid in the construction of NS-SEC – a measure that will be used in subsequent sensitivity analysis – makes it extremely attractive to conduct a sensitivity analysis of SOC codes to compare SOC 90 and SOC 2000 codes both within and between different social stratification measures.

The major differences between SOC 90 and SOC 2000 relate to managerial occupations and the addition of computing, technology, environment and conservation, and customer service occupations (*SOC 2000 - Office for National Statistics*, 2000). Whilst the same rough breakdown of major groups, sub-major groups, minor groups, and unit groups is kept between SOC 90 and SOC 2000 there are significant differences, one of which being the number of categories within each as seen in table 2.6.

Table 2.5 Breakdown of classification of SOC 90 and SOC 2000

|  |  |  |
| --- | --- | --- |
| Groups | SOC 90 | SOC 2000 |
| Major | 9 | 9 |
| Sub-major | 22 | 25 |
| Minor | 77 | 81 |
| Unit | 371 | 353 |

Note: Table taken from (*SOC 2000 - Office for National Statistics*, 2000)

All these changes related to the introduction of new occupations, de-industrialisation, and a desire to make SOC more compatible with ISCO meant that SOC 90 and SOC 2000 do have some discontinuities as seen by a side-by-side comparison of sub-major groupings attributed by skill level in table 2.7.

Table 2.6 Sub-major groups of SOC 90 and SOC 2000 by Skill Level

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Skill Level | SOC 90 | | SOC 2000 | |
| Level 4 | 1a | Corporate managers and administrators | 11 | Corporate managers |
| 2a | Science and engineering professionals | 21 | Science and technology professionals |
| 2b | Health professionals | 22 | Health professionals |
| 2c | Teaching professionals | 23 | Teaching and research professionals |
| 2d | Other professional occupations | 24 | Business and public service professionals |
| Level 3 | 1b | Mangers/proprietors in agriculture and services | 12 | Managers and proprietors in agriculture and services |
| 3a | Science and engineering associate professionals | 31 | Science and technology associate professionals |
| 3b | Health associate professionals | 32 | Health and social welfare associate professionals |
| 6a | Protective service occupations | 33 | Protective service occupations |
| 3c | Other associate professional occupations | 34 | Culture, media and sports occupations |
| 7a | Buyers, brokers and sales representatives | 35 | Business and public service associate professionals |
| 9a | Other occupations in agriculture, forestry and fishing | 51 | Skilled agricultural trades |
| 5b | Skilled engineering trades | 52 | Skilled metal and electrical trades |
| 5a | Skilled construction trades | 53 | Skilled construction and building trades |
| 5c | Other skilled trades | 54 | Textiles, printing and other skilled trades |
| Level 2 | 4a | Clerical occupations | 41 | Administrative occupations |
| 4b | Secretarial occupations | 42 | Secretarial and related occupations |
| 6b | Personal service occupations | 61 | Caring personal service occupations |
| 7b | Other sales occupations | 62 | Leisure and other personal service occupations |
| 8a | Industrial plant and machine operators, assemblers | 71 | Sales occupations |
| 8b | Drivers and mobile machine operators | 72 | Customer service occupations |
| 81 | Process, plant and routine operatives |
| 82 | Transport and mobile machine drivers and operatives |
| Level 1 | 9b | Other elementary occupations | 91 | Elementary trades, plant and storage related occupations |
| 92 | Elementary administrative and service occupations |

Note: Table taken from (Mackinnon, 2001)

Within subsequent analysis, alongside a sensitivity analysis of social stratification measures, another sensitivity analysis will be conducted comparing SOC 90 and SOC 2000 codes. Both the SOC 90 and SOC 2000 construction of each of the three social stratification variables: NS-SEC, CAMSIS, and RGSC will be provided in the following descriptive statistics. Goodness-of-fit statistics alongside theoretical obligations will determine the best fit social stratification measure to use within subsequent analysis, as such the comparative SOC 90 and SOC 2000 models will only be compared for the selected social stratification variable – the other two comparisons will be provided in the appendix[[13]](#footnote-13).

##### Registrar General Class Schema

The Registrar General’s Social Class is a social class measures that originated in the early 20th century in the UK – first used in 1911 to show variation in infant mortality according to parents’ occupation (Stevenson, 1913). This measure of social stratification was later re-developed in 1921 and again in 1928 by stating that class was more closely equated with occupation than material factors of income or wealth in explaining certain phenomena[[14]](#footnote-14) (Stevenson, 1928). The measure is built upon the assumption that society is graded based on a hierarchy of occupations (Murray 2011). The original conceptualisation of the RGSC schema was based upon the assumption that unit groups were allocated to a given social class whereby said classification was commensurate with the degree of expertise involved in carrying out the jobs of that occupation (Rose and Pevalin, 2001). From 1921 to 1971 the RGSC schema was an ordinal classification that was based upon individuals ‘standing within their community’ (ibid). This ‘standing within their community’ was replaced by a new definition, one that was based upon occupational skill (ibid). Though occupational skill had always had some level of relevance within the RGSC this was the first time it was made an explicit definition of the RGSC (Prandy, 1990). This demonstrates a departure from a stricter Weberian definition of status groups and towards a more traditional understanding of social class (ibid).

The RGSC rests upon a theoretical assumption that social inequality exists within society and that individuals are socially stratified by unequal rewards (Szreter, 1984). This social inequality is structured around a single scale of social position/status within society encapsulated within occupational categories. These occupational categories form a single uni-dimensional hierarchy across all of Britain. The original creator of the schema, Stevenson, created the model of RGSC based upon an assumption that society comprises an upper-middle, middle, and working class (Prandy, 1999). This assumption is baked into the theoretical implications of the unidimensional hierarchy. The RGSC schema also follows an explicit hierarchical ordering split into two halves: a non-manual dimension at the top half of the scheme and a manual dimension at the bottom half of the scheme, as seen in Table 1.5.

The schema is broken into six categories, from unskilled manual occupations to higher-level professionals (ibid). The RGSC once formed the basis of all commonly used social classifications within Britain (Szreter 1984). Alternative measures like the National Statistics Socio-Economic Classification have become prominent. This is mostly due to the building criticism of the RGSC schema. These critiques state that the RGSC schema is an a priori scale (Rose and Pevalin, 2001) that in its 50 or so years of use has never been externally validated (Prandy, 1990). Attempts to validate RGSC by the likes of Bland (Bland, 1979) using the Hope-Goldthorpe scale (Goldthorpe and Hope, 1974) have raised serious doubts about the use of not only the manual/non-manual divide but also the coding structure of the RGSC (Bland, 1979) – Bland argued that over a third of the schema needs to be re-coded due to a lack of validation of the schema and its original manifestation having no concrete division between skilled non-manual and skilled manual occupations (Prandy, 1990). This presents a view that compared to other more robust and externally validated measures the RGSC will produce alternative substantive interpretations to those that have been robustly measured and coded. The strong views from the likes of Bland (1979) suggest that the RGSC class schema is not an adequate sociological measure of social class. This provides another reason to produce a sensitivity analysis using different social stratification measures.

Acknowledging the apparent critiques of the RGSC, the measure itself has been used within social stratification research and is often included as a measure of social class in datasets (Gregg, 2012). The RGSC measure existed for 47 years before the NCDS existed. Compared to other social stratification measures, such as NS-SEC, created after the NCDS, this presents an opportunity to test the substantive interpretations of different social stratification measures constructed at different times. This temporal distinction will become an essential part of the subsequent sensitivity analyses.

The Full RGSC class schema is detailed below in table 2.8:

Table 2.7 RGSC Class Schema

|  |  |  |  |
| --- | --- | --- | --- |
|  | Class | Occupations | Example Occupations |
| Non-Manual | I | Professional Occupations | Accountant |
| II | Intermediate Occupations | Police Officer |
| IIIN | Skilled Non-Manual Occupations | Clerical Worker |
| Manual | IIIM | Skilled Manual Occupations | Butcher |
| IV | Partly Skilled Occupations | Postal Worker |
| V | Unskilled Occupations | Labourer |

Amongst the social stratification variables that are provided for the NCDS and BCS, full-auto, semi-auto, and verification processing variables are provided. These are different methods for coding the entire batch of occupational codes into a given social stratification variable. The automatic method uses a computer to automatically place a given observation in their most likely occupational code. The semi-automatic method uses a threshold based on a score of confidence for how likely it is true that an observation really does fall into that given occupational code, for the semi-automatic method a threshold of 45 out of 100 was used, any observation below that threshold was then manually coded by a researcher and verified by another (Gregg, 2012). Semi-auto processing social stratification variables are used within subsequent analysis as suggested (Gregg, 2012). Due to both SOC2000 and SOC90 codes being provided, both will be used, and a sensitivity analysis will be conducted to assess the best fit model alongside a sensitivity analysis of social stratification measures.

The UKHLS provides SOC codes for… [More added and completed when UKHLS data is available].

##### National Statistics Socio-Economic Classification

Rose and Pevalin developed the NS-SEC schema (Rose and Pevalin, 2002). The operational categories of the NS-SEC represent labour market positions, employment statuses, and employment relations. The production of NS-SEC as a class schema in part came from the rising critiques of former measures such as the RGSC. Goldthorpe originally argued for a new measure of social class where “a measure of class will be most apt where the link to the dependent variable is believed theoretically to be through the individual's position in relations of production; a measure of status... where the link is believed to be through positions in relations of consumption or literature” (Goldthrope, cited in Rose and Pevalin, 2001: 13). This gave birth to the EGP social class measure. The Office for National Statistics (ONS) and the Economic Social Research Council (ERSC) commissioned a further social class measure on the theoretical basis of Goldthorpe’s position, which gave birth to the NS-SEC class schema.

NS-SEC was developed from the Erikson-Goldthorpe-Portocareo (EGP) perspective (Rose and Pevalin, 2002). The EGP scheme, like NS-SEC, rested on theoretical assumptions of labour market positions, employment status, and relations and was initially developed by Goldthorpe (Goldthorpe, 1980).

Employment relations are central to the NS-SECs' ideas on social class – and the development of social class schemas. These employment relations are split into three distinctive formations: those that purchase labour and have authority over those they have purchased labour from, self-employed workers, and employees who sell their labour and are thus under the authority of employers (Erikson, Goldthorpe and Portocarero, 1979, 1982, 1983). This differentiation of employment relations gives rise to class-based patterns of social stratification (Williams, 2017).

Like other social class schemas already mentioned, a central tendency for the EGP and eventually the NS-SEC study of social class rests upon an analysis of relationships – one occupational group is relational to another within the broader social class schema (Goldthorpe and Marshall, 1992). One major difference between the former RGSC schema and the NS-SEC schema was a fundamental rejection of the manual/non-manual divide (ISER, 2024). NS-SEC rejects the idea that there is such a divide in so far as that means a broad division between the middle and working classes – with some manual occupations occupying classes in Class 3 and other non-manual occupations in Class 6 or 7 (ibid).

The complete NS-SEC classification schema has 14 operational categories related to employment relations but can be broken down into as few as three analytical categories. For this analysis NS-SEC is broken down into its nine analytical class variety (Rose and Pevalin, 2010).

Table 2.8 NS-SEC Class Schema

|  |  |
| --- | --- |
|  | Analytical Variables for NS-SEC |
| Operational Categories |  |
| L1  Employers in large establishments | 1.1 Large Employers and higher managerial occupations |
| L2  Higher managerial occupations |
| L3  Higher professional occupations | 1.2 Higher professional occupations |
| L4  Lower professional and higher technical occupations | 2 Lower Managerial and professional occupations |
| L5  Lower managerial occupations |
| L6  Higher supervisory occupations |
| L7  Intermediate occupations | 3 Intermediate occupations |
| L8  Employers in small establishments | 4 Small employers and own account workers |
| L9  Own account workers |
| L10  Lower supervisory occupations | 5 Lower supervisory and technical occupations |
| L11  Lower technical occupations |
| L12  Semi-routine occupations | 6 Semi-routine occupations |
| L13  Routine occupations | 7 Routine occupations |
| L14  Never worked and long-term unemployed | 8 Never worked and long-term unemployed |

The following analytical variables within the NS-SEC have been broken down with example occupations to aid in interpretation within subsequent models in Table 2.10.

Table 2.9 Examples of Occupations from Analytical NS-SEC

|  |  |
| --- | --- |
| Analytical Variables for NS-SEC | Example Occupations |
| 1.1 Large Employers and higher managerial occupations | Chief Executives, Managers and directors in finance |
| 1.2 Higher professional occupations | Lecturers, Judges, Doctors |
| 2 Lower Managerial and professional occupations | Managers in retail, Nurses, School Teachers |
| 3 Intermediate occupations | Paramedics, Teaching Assistants |
| 4 Small employers and own account workers | Bricklayers, Carpenters, Shopkeepers |
| 5 Lower supervisory and technical occupations | Mechanics, Plumbers, Skilled Construction Supervisors |
| 6 Semi-routine occupations | Sales Assistants, Veterinary Nurses |
| 7 Routine occupations | Cleaners, Welding Trades, Hairdressers |

##### CAMSIS

The Cambridge Social Interaction and Stratification Scale (CAMSIS) represents a social stratification scale based on measures of relative social distance (Prandy and Lambert, 2003). These relationship networks are ultimately hierarchical and reify themselves in reproducing hierarchical inequalities (Bergman and Joye, 2001).

The continuous nature of CAMSIS means that numerical values are attached to occupations, meaning the relative value of each occupational value is only meaningful compared to other occupations on the same scale (Connelly et al. 2016). This is meaningful when it comes to the interpretation of the CAMSIS measure within models of analysis, as the value of the coefficient is always going to be concerning the comparison to other occupations along the CAMSIS scale. The most considerable difference between CAMSIS and other social stratification measures discussed is that CAMSIS does not believe that distinct groups in the form of social classes are differentiated through material and status differences. CAMSIS rejects a ‘simple structuralism’ dependent on a static structure – it proposes a much more dynamic, constantly re-constitutive process (Bergman and Joye, 2001).

While CAMSIS stands in contrast to the other social stratification measures mentioned, they share some similarities. CAMSIS contends - as do the NS-SEC and RGSC – that occupational groups are the primary mechanism by which social and economic rewards are distributed within modern societies (ibid) and, as such, are some of the best indicators of social stratification in society.

The full CAMSIS scale typically has a mean of around 50 and a standard deviation of around 15. Occupations such as cleaners would find themselves at the lower end of this scale, with occupations like judges being around 86. Whilst a table of CAMSIS scores would have to include all SOC codes multiplied by two (for men and women) and therefore be too long to include here, a comparison using some select SOC codes can be made to compare with RGSC and NS-SEC tables.

Table 2.10 Examples of CAMSIS scores by SOC-90 Codes

|  |  |  |  |
| --- | --- | --- | --- |
| SOC-90 | SOC-90 Label | Male CAMSIS SOC90 | Female CAMSIS SOC90 |
| 100 | General administrators; national government | 85.6 | 67.9 |
| 231 | Higher and Further education teaching professionals | 63.8 | 78.6 |
| 450 | Medical secretaries | 62.6 | 65.8 |
| 733 | Scrap dealers, scrap metal merchants | 44 | 42.9 |
| 950 | Hospital porters | 38.5 | 50.7 |

Due to the CAMSIS scale used in this analysis does not conform to the distribution expected, this is most likely because the simplified CAMSIS scale is used as data availability restricts the use of the full-implementation of CAMSIS. The expected mean of 50, s.t.d of 15 is typically found in the full-implementation of CAMSIS.

## Modelling youth’s first major transition – a pooled analysis

Each cohort and dataset have been described in detail and important socio-historical context has been provided. Past empirical literature has been highlighted and important social theory has been earmarked for detailed discussion after substantive interpretation of analyses. This chapter seeks to conduct a proposed logistic regression model using economic activity as a binary dependent variable to assess the strength and size of substantive effects structural inequalities have on young peoples first major transition post-mandatory schooling. This analysis will pool data[[15]](#footnote-15) from all datasets – the NCDS, BCS, BHPS, and UKHLS samples. This analysis will study within cohort effects of structural inequalities as well as the between cohort differences -- if any – that are present amongst Britain’s youth. This chapter will first provide a breakdown of the descriptive statistics of the proposed complete records model for analysis. This proposed analysis will use previously described independent variables associated with structural inequality – educational attainment, sex, housing tenure, and social class[[16]](#footnote-16). After detailing the descriptive story of the pooled sample, and by cohort sample, analysis will proceed using a logistic regression model. This chapter will conclude with a discussion on the substantive story within and between each cohort and how that links to prior discussions on past empirical and theoretical literature on the subject of the sociology of youth.

The logistic regression model output uses the reference category of 'Don't Continue Schooling'. The reference category contrasts with continuing schooling. The continue schooling versus not continue schooling is sociologically compelling. Contrasting continuing school with other economic activity destinations is temporally relevant given the possible impact that increasing the mandatory school leaving age, decline in the manufacturing industry, and rise in part-time work may have on the economic destinations of youth. Less than five O’levels is the reference category for educational attainment, Female is the reference category for Sex, Own home is the reference category for housing tenure, and NS-SEC 2 is the reference category for NS-SEC.

#### Descriptive Statistics

Table 2.10 details the descriptive statistics for the following model. These descriptive statistics have a total N=10,039. This covers two cohorts, the NCDS and BCS. The NCDS has a n=8,411 and the BCS has a n=1,628. Some statistics will now be discussed.

Table 2.11 Pooled Cohort Descriptive Statistics

|  |  |  |
| --- | --- | --- |
| Table 2.10: Descriptive Statistics for Economic Activity (Combined Model) | | |
|  | n | % |
| Economic Activity |  |  |
| Don't Continue Schooling | 4,099 | 41.05% |
| Continue Schooling | 5,886 | 58.95% |
| Cohort |  |  |
| NCDS | 6,387 | 63.97% |
| BCS | 3,598 | 36.03% |
| Educational Attainment O'levels |  |  |
| Less than Five O’levels | 5,087 | 50.95% |
| Five or More O’levels | 4,898 | 49.05% |
| Sex of Respondent |  |  |
| Female | 5,245 | 52.53% |
| Male | 4,740 | 47.47% |
| Housing Tenure of Respondent when a Child |  |  |
| Own Home | 350 | 3.51% |
| Don't Own Home | 528 | 5.29% |
| Semi-Dominant NS-SEC Social Class of Parents when Respondent was 10 SOC2000 | 1,334 | 13.36% |
| Large Employers and higher managerial occupations | 984 | 9.85% |
| Higher professional occupations | 1,194 | 11.96% |
| Lower Managerial and professional occupations | 1,630 | 16.32% |
| Intermediate occupations | 1,683 | 16.86% |
| Small employers and own account workers | 2,282 | 22.85% |
| Lower supervisory and technical occupations |  |  |
| Semi-routine occupations | 8,411 | 84.24% |
| Routine occupations | 1,574 | 15.76% |
|  |  |  |
| n |  | 9985 |
| Data Source: NCDS & BCS | | |

Table 2.12 Pooled descriptive statistics by Cohort

|  |  |  |  |
| --- | --- | --- | --- |
| Descriptive Statistics by Cohort | | | |
|  | Cohort | | |
|  | NCDS | BCS | Total |
| n | 8411 (83.78%) | 1574 (15.76%) | 9985 (100.00%) |
| Economic |  |  |  |
| Don't Continue Schooling | 5116 (60.83%) | 624 (39.64%) | 4099 (41.05%) |
| Continue Schooling | 3295 (39.17%) | 950 (60.36%) | 5886 (58.95%) |
| Educational Attainment O'levels |  |  |  |
| Less than Five O’levels | 5426 (64.51%) | 961 (61.05%) | 6387 (63.97%) |
| Five or More O’levels | 2985 (35.49%) | 613 (38.95%) | 3598 (36.03%) |
| Sex of Respondent |  |  |  |
| Female | 4215 (50.11%) | 872 (55.40%) | 5087 (50.95%) |
| Male | 4196 (49.89%) | 702 (44.60%) | 4898 (49.05%) |
| Housing Tenure of Respondent when a Child |  |  |  |
| Own Home | 4045 (48.09%) | 1200 (76.24%) | 5245 (52.53%) |
| Don't Own Home | 4366 (51.91%) | 374 (23.76%) | 4740 (47.47%) |
| Semi-Dominant NS-SEC Social Class of Parents when Respondent was 10 SOC2000 |  |  |  |
| Large Employers and higher managerial occupations | 261 (3.10%) | 89 (5.65%) | 350 (3.51%) |
| Higher professional occupations | 410 (4.87%) | 118 (7.50%) | 528 (5.29%) |
| Lower Managerial and professional occupations | 1038 (12.34%) | 296 (18.81%) | 1334 (13.36%) |
| Intermediate occupations | 805 (9.57%) | 179 (11.37%) | 984 (9.85%) |
| Small employers and own account workers | 1024 (12.17%) | 170 (10.80%) | 1194 (11.96%) |
| Lower supervisory and technical occupations | 1372 (16.31%) | 258 (16.39%) | 1630 (16.32%) |
| Semi-routine occupations | 1485 (17.66%) | 198 (12.58%) | 1683 (16.86%) |
| Routine occupations | 2016 (23.97%) | 266 (16.90%) | 2282 (22.85%) |

### Modelling Youth’s First Major Transition

Each analytical variable within the model is presented with its individual cohort level effect. For example, the educational attainment impact is assessed at the NCDS and BCS levels. A cohort analytical variable is also included in analysis to assess the cohort level impact – NCDS is used as the reference cohort category. Adding the cohort level coefficient to the BCS level individual effects produces the cohort interaction effect. Doing this with log odds is intuitively difficult, as such average marginal effects are provided for easier interpretation of between cohort level effects.

Logistic regression results can be interpreted using a variety of measures; log odds, relative risk ratios, odds ratios, and marginal effects to name the most common. When dealing with logistic regression, results in the form of coefficients are reported in the default Stata output as log odds. Log odds are notoriously tricky to interpret and are rarely well described in sociological studies (Gayle and Lambert, 2009). For example, for a categorical explanatory variable, the coefficient associated with category effects is considered the effect on the log odds of moving from the reference category to the category of the X variable. Due to this difficulty in interpretation, some (Norton and Dowd, 2018) have advocated for using odds ratios over log odds. Odds ratios are relatively simple to calculate and applicable to both continuous and discrete explanatory variables of interest.

However, odds ratios lack an intuitive interpretation, which results in an inability to compare across models and datasets, even if they have the exact model specification (ibid). An odds ratio is not an absolute number, it is conditional on the sample and on model specification (ibid). This issue stems from odds ratios changing if variables are added to the model, even if such additional variables are independent of the other variables – this is also true of log odds due to the fixed variance problem. Unlike linear regression, logistic regression coefficients don’t have an intuitive interpretation. Log odds are arguably beneficial so long as researchers interpret the substantive effects as ‘higher’ or ‘lower’ (Gayle and Lambert, 2009).

A possible alternative to using logs odds and odds ratios is the average marginal effect of an explanatory variable on the probability that equals 1 versus 0. In the case of this model, the average change in probability of continuing schooling over not continuing schooling, holding all other variables at their observed values. The rationale for interpreting non-parametric models using average marginal effects is because the marginal effect is interpreted as a percentage point probability on Y=1 making it a more intuitive alternative to something like an odds ratio. In addition, the average marginal effect can be either positive or negative.

Following the suggestions made by Connelly, Gayle, and Lambert (2016b), all non-linear models reported henceforth will include log odds to provide an estimate of the direction of the variable effect, average marginal effects to translate this into percentage point change, and where possible the inclusion of quasi-variance statistics and quasi-variance standard errors.

Log odds and average marginal effects are present in table 2.12. The between level average marginal effects are presented by subtracting the NCDS effect from the BCS effect. Following this a Cohort level effect is provided and added to the BCS level effects to provide a direct cohort comparison.

Table 2.13 Modelling First Major Transition with Combined Cohorts

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Pooled Model | | | Average Marginal Effects | |
| Economic Activity: ‘Don’t Continue Schooling’ Reference Category | **Log Odds** | **S.E.** | **Sig.** | **Prob.** | **S.E.** |
| Continue Schooling |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |
| *Less than five O’levels* | Ref. | Ref. |  |  |  |
| *Five or More O’levels # NCDS* | 2.98 | (0.06) | \*\*\* | 0.55 | (0.01) |
| *Five or More O’levels # BCS* | 1.21 | (0.12) | \*\*\* | 0.19 | (0.02) |
| Sex |  |  |  |  |  |
| *Female* | Ref. | Ref. |  |  |  |
| *Male # NCDS* | -0.50 | (0.06) | \*\*\* | -0.07 | (0.01) |
| *Male # BCS* | -0.58 | (0.11) | \*\*\* | -0.08 | (0.01) |
| Housing Tenure |  |  |  |  |  |
| *Own Home* | Ref. | Ref. |  |  |  |
| *Do not Own Home # NCDS* | -0.63 | (0.06) | \*\*\* | -0.09 | (0.01) |
| *Do not Own Home # BCS* | -0.20 | (0.13) |  | -0.03 | (0.02) |
| NS-SEC (SOC 2000) |  |  |  |  |  |
| *1.1 # NCDS* | 0.02 | (0.19) |  | 0.00 | (0.03) |
| *1.1 # BCS* | 0.29 | (0.29) |  | 0.04 | (0.04) |
| *1.2 # NCDS* | 0.48 | (0.17) | \*\* | 0.08 | (0.03) |
| *1.2 # BCS* | 0.12 | (0.26) |  | 0.02 | (0.04) |
| *2* | Ref. | Ref. |  |  |  |
| *3 # NCDS* | -0.25 | (0.13) | \* | -0.04 | (0.02) |
| *3 # BCS* | -0.20 | (0.21) |  | -0.03 | (0.03) |
| *4 # NCDS* | -0.89 | (0.12) | \*\*\* | -0.14 | (0.02) |
| *4 # BCS* | -0.57 | (0.21) | \*\* | -0.08 | (0.03) |
| *5 # NCDS* | -0.76 | (0.11) | \*\*\* | -0.12 | (0.02) |
| *5 # BCS* | -0.70 | (0.19) | \*\*\* | -0.09 | (0.02) |
| *6 # NCDS* | -0.89 | (0.11) | \*\*\* | -0.14 | (0.02) |
| *6 # BCS* | -0.35 | (0.20) |  | -0.05 | (0.03) |
| *7 # NCDS* | -1.11 | (0.11) | \*\*\* | -0.17 | (0.02) |
| *7 # BCS* | -0.50 | (0.19) | \*\* | -0.07 | (0.03) |
| Cohort |  |  |  |  |  |
| *NCDS* | (.) | (.) |  |  |  |
| *BCS* | 1.03 | (0.18) | \*\*\* | 0.16 | (0.03) |
| Intercept | -0.40 | (0.09) | \*\*\* |  |  |
| Number of Observations | 9985 | | |  |  |
| McFadden’s | 0.35 | | |  |  |
| McFadden’s Adjusted Pseudo | 0.34 | | |  |  |
| Cox-Snell Pseudo | 0.38 | | |  |  |
| Nagelkerke Pseudo | 0.51 | | |  |  |
| Tjur’s | 0.42 | | |  |  |
| AIC | 8954.76 | | |  |  |
| BIC | 9113.35 | | |  |  |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: NCDS [Sweeps 0-4] & BCS [Sweeps 0-5] | | | | | |

A graph with black and white lines

Description automatically generated

Figure ‎2.2 Log Odds Coefficient Plot of Pooled Logistic Regression Analysis of Youth's First Transition

Figure 2.2 provides a graphical visualisation of the log odds coefficients within the analysis; statistics are provided in table 2.12. For members of the NCDS, receiving five or more O’levels compared to receiving less than five O’levels presents an increased log odds of continuing schooling. Translated into average marginal effects, this results in a 55 percentage point increase of continuing schooling. For members of the BCS, there is also an increased log odds of continuing schooling, or a 19 percentage point increase of continuing schooling. There is a 36 percentage point difference on the impact of educational attainment in continuing schooling between members of the NCDS and BCS cohort. The members of the BCS cohort in of itself has an increased log odds of continuing schooling over that of members of the NCDS cohort, translated to average marginal effects this is a 16 percentage point increase of continuing schooling. Added to the BCS educational attainment impact, this results in members of the BCS that received five or more O’levels having a 35 percentage point increased probability of continuing schooling, reducing the overall difference between cohorts from 36 percentage points to 19 percentage points. This still represents a substantively significant difference in the scale of impact educational attainment has on individuals of different cohorts. Both cohorts exhibit educational attainment as the single largest effect on young peoples first major transition. Educational attainment is a major barrier to continuing schooling for both NCDS and BCS youth. However, the impact of educational attainment appears to have diminished substantive across cohorts. This is suggestive of a growing demand for continuing education beyond the barrier of prior educational attainment. This can be explained by the formation of a credentialled labour market – requiring certifications and degrees for occupations that previously did not require them. This would entirely explain the decreased impact of education between the NCDS and BCS cohorts.

Figure 2.3 details the predictive margins of educational attainment by cohort. Whilst both the NCDS and BCS cohort share a similar predictive margin of continuing education if an individual attains less than five O’levels, divergences appear if individuals attain five or more O’levels. The gradient comparing educational attainment of the NCDS cohort is much steeper than it is for the BCS cohort, even though the same general trend applies to both cohorts. This appears indicative of a declining impact of educational attainment upon continuing schooling for youth.

A screenshot of a computer

Description automatically generated

Figure ‎2.3 Predictive Margins of Educational Attainment by Cohort

Moving on to Sex, for members of the NCDS, being a man compared to being a woman presents a decreased log odds of continuing schooling. Translated to average marginal effects this represents a 7 percentage point decrease of continuing schooling if a man in the NCDS compared to a woman. The BCS follows a near identical pattern, with an 8 percentage point decrease of continuing schooling if a man compared to a woman. There is a slight increased penalty by 1 percentage point of being a man in the BCS cohort compared to being a man in the NCDS cohort. However, when adding the cohort specific effect, for men in the BCS cohort there is an 8 percentage point increase of continuing schooling, the difference between men in the NCDS cohort and men in the BCS cohort thus becomes 15 percentage point. Once more, a substantively significant difference primarily produced via cohort level change. The lack of substantive difference between the cohorts suggests that sexed based differences within British society are – at least for the timeframe from the NCDS to BCS – rigid. The sex based influences that decrease the likelihood of men continuing schooling compared to women appear to extent above and beyond one cohorts socio-historical context and provide a period based effect.

Figure 2.4 depicts the predictive probabilities of sex by cohort difference. The predictive margins display contrasting findings by cohort. For the NCDS women have a lower predictive margin of staying on within schooling compared to their male counterparts – which makes sense for their particular socio-historical context. For the BCS cohort on the other hand, women have a larger predictive margin of continuing schooling compared to men – this is perhaps reflective of the increasing educational participation of women in the labour market and exogenous pressures on women to stay on within education. It is important to reflect back to table 2.13 and remember that whilst these predictive margins are contrasting, the average marginal effects and log odds for sex for both the NCDS and BCS are near identical and show the same overall trend – men are less likely than women to continue schooling.

A screenshot of a computer

Description automatically generated

Figure ‎2.4 Predictive Margins of Sex by Cohort

Moving on to housing tenure, for members of the NCDS, living in a home not owned by their parents presents a decreased log odds of continuing schooling compared to individuals that had parents that owned their own home. Translated into average marginal effects relates to a 9 percentage point decrease of continuing schooling. For the BCS housing tenure is not statistically significant and so no effect is interpreted. The lack of statistical significance implies a relative decline in the impact and influence of housing tenure over individuals’ school-to-work transitions.

Figure 2.5 displays the predictive margins for housing tenure by cohort. Whilst the NCDS cohort demonstrates that owning your own home has a higher predictive margin to continue schooling than not owning your own home, the BCS cohort demonstrates a lack of statistical significance due to overlapping confidence intervals.

A screenshot of a computer

Description automatically generated

Figure ‎2.5 Predictive Margins of Housing Tenure by Cohort

Finally moving on to NS-SEC. Due to its categorical nature, NS-SEC will be interpreted comparing like for like categories across each cohort specific effect, with general trends discussed at the end. For members of the NCDS whose parents are in NS-SEC 1.1 compared to NS-SEC 2

The cohort variable on its own suggests that there is an increased log odds of continuing schooling for individuals in the BCS cohort compared to those in the NCDS cohort. Translated to average marginal effects, there is a 16 percentage point increase of continuing schooling if individuals are in the BCS cohort compared to the NCDS cohort.

For members of the NCDS cohort, individuals whose parents are in NS-SEC 4 have a decreased log odds of continuing schooling, or a 14 per cent decreased probability of continuing schooling compared to NS-SEC 2. For members of the BCS cohort, individuals whose parents are in NS-SEC 4 also have a decreased log odds or an 8 per cent decreased probability of continuing schooling. There is an increase of 6 per cent between NCDS to BCS cohorts for NS-SEC 4 members, demonstrating that whilst NS-SEC 4 members compared to NS-SEC 2 maintain a negative probability of continuing schooling across cohorts, there is a decline in the strength of this probability from NCDS to BCS cohorts. When adding the cohort specific effect to this phenomena, even though members of NS-SEC 4 within the BCS have an 8 per cent decreased probability of continuing schooling compared to NS-SEC 2, being members of the BCS means that NS-SEC 4 members have an 8 per cent increased probability of continuing schooling compared to NCDS members.

For members of the NCDS cohort, individuals whose parents are in NS-SEC 5 have a decreased log odds of continuing schooling, or a 12 per cent decreased probability of continuing schooling compared to NS-SEC 2. For members of the BCS cohort, individuals whose parents are in NS-SEC 5 also have a decreased log odds or a 9 per cent decreased probability of continuing schooling. There is an increase of 3 per cent between NCDS to BCS cohorts for NS-SEC 5 members, demonstrating that whilst NS-SEC 5 members compared to NS-SEC 2 maintain a negative probability of continuing schooling across cohorts, there is a decline in the strength of this probability from NCDS to BCS cohorts. When adding the cohort specific effect to this phenomena, even though members of NS-SEC 5 within the BCS have a 9 per cent decreased probability of continuing schooling compared to NS-SEC 2, being members of the BCS means that NS-SEC 4 members have a 7 per cent increased probability of continuing schooling compared to NCDS members.

For members of the NCDS cohort, individuals whose parents are in NS-SEC 6 have a decreased log odds or 14 per cent decreased probability of continuing schooling compared to their NS-SEC 2 peers. For the BCS cohort NS-SEC 5 is not statistically significant and implies a declining influence in social origins for NS-SEC 5 across cohorts.

Finally, for members of the NCDS cohort, individuals whose parents are in NS-SEC 7 have a decreased log odds of continuing schooling, or a 17 per cent decreased probability of continuing schooling compared to NS-SEC 2. For members of the BCS cohort, individuals whose parents are in NS-SEC 7 also have a decreased log odds or a 7 per cent decreased probability of continuing schooling. There is an increase of 10 per cent between NCDS to BCS cohorts for NS-SEC 7 members, demonstrating that whilst NS-SEC 7 members compared to NS-SEC 2 maintain a negative probability of continuing schooling across cohorts, there is a decline in the strength of this probability from NCDS to BCS cohorts. When adding the cohort specific effect to this phenomena, even though members of NS-SEC 7 within the BCS have a 7 per cent decreased probability of continuing schooling compared to NS-SEC 2, being members of the BCS means that NS-SEC 7 members have a 9 per cent increased probability of continuing schooling compared to NCDS members.

Graphical visualisations have been produced to demonstrate the predicted probabilities and average marginal effects of NS-SEC within the model in figure 2.2. The graphed average marginal effects help to demonstrate the similarities between cohorts for the impact NS-SEC has upon continuing schooling. Both the average marginal effects and predictive margins displayed in figure 2.2 demonstrate a shared trend between cohorts for NS-SEC – though there are some minor divergences related to the predictive probabilities of NS-SEC 4 and 5 between the cohorts.

A screenshot of a graph

Description automatically generated

Figure ‎2.6 AMEs and Predictive Margins by Cohort

Both members of the NCDS and BCS follow similar general trends with reference to NS-SEC. Both share an overall pattern whereby those members of the lower half of NS-SEC have the largest decreased probability of continuing schooling compared to NS-SEC 2. Where they differ is in the size of these effects – the NCDS cohort has a greater effect size compared to the BCS cohort – and in the nature of statistical significance in particular in reference to the upper half of NS-SEC.

The general substantive conclusion from this model presents a similar analytical trend in the substantive results between the NCDS and BCS cohort – though the BCS cohort appears to present more muted results compared to the NCDS cohort. What is substantive, is the impact of the cohort specific effect. For the BCS cohort, whilst it maintains a similar effect trend compared to the NCDS cohort, the cohort specific effect added to such effects in almost all cases changes the direction of the effect – whether that be in terms of Sex or NS-SEC. Even though members of the BCS cohort share a similar negative probability of continuing schooling, the cohort effect makes it so that men whilst having a near identical negative probability to continue schooling as their NCDS counterparts, by virtue of being in the BCS cohort they actually have an increased probability of containing schooling compared directly with the NCDS cohort. Whilst there appears to maintain across cohorts a significant Sex and social class effect for example, these effects are improved comparative to previous cohorts.

### Discussion and Conclusions

The provision of repeated contacts based data has enabled an analysis of the life course within the life domain of school-to-work transitions for youth (Diewald and Mayer, 2008; Mayer, 2009). This repeated contacts data provides insight into the role of structural inequalities on the sorting of youth into continuation of schooling. Focusing upon a particular life domain of school-to-work, repeated contacts data has demonstrated that structural inequalities have an impact upon youths’ transitionary experiences.

A core argument of Individualisation theory proposes that over time, structures weaken and ‘detraditionalize’ as individuals have to rely upon their own self (Beck, 1992). Agents navigate society that is increasingly managed and organised by individualised lifeworld. A comparison of the NCDS and BCS cohort models presented in table 2.13 provide some evidence for this thesis. The strength of educational attainment has weakened between the NCDS to BCS period, as has social class. Housing tenure has weakened to the point of not being statistically significant at all between the NCDS and BCS cohorts. Only sex has maintained its strength across cohorts. The evidence of a process of de-traditionalization found. A weakening of social structures is also found. Evidence points to a level of degradation in structural influences upon youths first transition. Whilst all but one structural inequality indicator weakened across cohorts, the cohort specific variable indicates that the socio-historical context of the BCS encourages more young people to continue schooling than the NCDS counterparts. The structural effects appear to have been subsumed under a more general widespread cohort-based effect of continuing education. This provides evidence for an increasingly credentialled labour force that demands continuing education for most of its workforce.

The collapse in housing tenure as an effective stratifying variable within the BCS could be seen as a collapse in housing tenure as an effective stratifying vairbale within wider British society. This is supported from the consequences of Thatcher’s Housing Act (*Housing Act*, 1980) which gave five million tenants in England and Wales the Right to Buy their house from their local authority with a discount. There is also the potential that missingness is underplaying the role of housing tenure. Records suggest that the true homeownership levels and rates of owner occupiers in England and Wales were 65.5 per cent in 1986 (HomeOwners Alliance, 2012). The analysis for the BCS cohort reports a homeownership rate of 76 per cent, an inflation of 10 per cent. Going forward, handling missing data sections will explore this possibility.

New Structuralism argues along similar epistemic grounds to the Individualisation thesis argued that a decline in traditional structural inequalities such as social class and sex would open a rise in new social cleavages such as housing tenure. Housing tenure from the NCDS model to the BCS model is the only variable to move from being statistically significant in one cohort to non-significant in another. New social cleavages are not becoming the dominant forms of stratification as time progresses.

This analysis has demonstrated that the structures that govern life domains have undergone a metamorphosis. The rules and resources of the school-to-work life domain are differentiated between the NCDS and BCS period. The initial cohort-based effect demonstrates that individuals within the BCS cohort were more likely to continue schooling over individuals within the NCDS cohort. The time-space contexts that govern the rules and resources of society have changed from the NCDS to BCS cohorts resulting in an increased continuation of schooling for the latter cohort. The expansion of education following a societal shift of purpose in education as a point of economic investment followed from the Crowther and Newsom reports (*Crowther Report Volume I*, 1959; *Newsom Report*, 1963) has promoted a rise of credentialism (Jones, 2016) which has precipitated a rise in demand for qualifications prior to labour market entry.

Whilst the interaction between sex and cohorts was found to not be statistically significant it would be too strong to argue this means there has been little change in the rules and resources that govern sex-based structures in British society between the NCDS and BCS cohorts. What is more likely is that the sex-based nuances of economic activity are lost to a simplified model that dichotomises choice into continuing versus not continuing schooling. If the model was expanded into multiple transition destination routes, sex-based differences should become more apparent between cohorts.

Overall, evidence has shown that theories related to New Structuralism are not empirically supported in this analysis – though there is some support for the Individualisation thesis, given that appropriate missing data mechanisms are checked for each sample pooled. All evidence points to a British society that has evolved the rules and resources that govern structures which in turn has altered slightly the effects structural inequalities have had upon individuals continuing schooling. This evidence makes a strong case that agency is situational and bounded to a particular space-time. The similarities between the NCDS and BCS suggest that certain structural inequalities are more engrained than others – such as social class and sex. Other analytical variables have seen a greater change in influence such as housing tenure and educational attainment. Both these variables appear to have had strong exogenous changes to the British environment that may have influenced their changing impact between the two cohorts. By analysing this empirical data through the lens of Structuralism and emphasising the bounded agency of individuals, it provides a compelling explanation to the statistical phenomena present in this chapter. Moving forward, a more detailed, granular analysis of each cohort will be conducted. Sensitivity analysis of social stratification measures and SOC codes will take place, as well as handling missing data sections to understand the potential bias missingness may have on the discussions and conclusions outlined above.

## In-depth NCDS Analysis

A pooled analysis of multiple cohorts provides a unique ability to assess between cohort effects in relation to the changing nature of structural inequalities impact upon youth’s first major transition. A pooled analysis does however limit the ability to provide an in-depth assessment of a particular dataset or sample. It is difficult for example, to assess different social stratification measures and their impacts on each dataset. It is also difficult to handle missing data with a pooled analysis. Therefore, to forgo these difficulties, each cohort sample will be assessed separately. This in-depth analysis will include a detailed breakdown of the sample, followed by a sensitivity analysis of social stratification measures and SOC codes as well as a handling missing data strategy to identify any potentially biased estimates from being reported. These detailed in-depth analyses of each individual cohort provide a more comprehensive overview of each individual cohort whilst providing greater between cohort information.

This section will focus on the National Childhood Development Study (NCDS). An initial logistic regression that follows the one presented in table 2.20 will be provided in greater detail. Following this, a sensitivity analysis of social stratification measures will be provided as well as an analysis using alternative standard occupation codes. A section on handling missing data will be provided – firstly a discussion of handling missing data strategies and a simulation of said strategies will be provided to select the best method going forward. Secondly, said method will be implemented within the NCDS analysis to assess missingness in the cohort analysis. Finally, a discussion and conclusions section will re-iterate the main findings in this section and provide any critique of the conclusions initially provided in the above section.

### Descriptive Statistics

Table 2.13 shows frequencies and summary statistics for the NCDS. Overall, 60.83 per cent of the sample is don’t continue schooling compared to 39.17 per cent that do continue schooling.

Regarding Educational Attainment, 64.51 per cent of individuals received less than 5 O’levels, with the remaining 35.49 per cent receiving five or more O’levels. Sex presents an equal split between men (48.09 per cent) and women (51.91 per cent). Regarding homeownership, 48.09 per cent of individuals grew up in a home owned by their parents compared to 51.91 per cent that did not.

The NS-SEC categories for SOC 2000 construction all see a relatively even distribution between 10-20 per cent except for the largest category – 7, at 23.97 per cent – and the smallest categories –1.1 and 1.2, at 3.10 per cent and 4.87 per cent. Compared to NS-SEC categories for SOC 90 construction, there are some substantive differences. Most prominent of these constitutes the exceedingly small number of individuals that occupy NS-SEC 1.1 – 0.11 per cent compared to the SOC 2000 constructions 3.10 per cent. Other differences see for the SOC 90 construction that NS-SEC 1.2, 2 and 4 decrease comparatives to their SOC 2000 counterparts and NS-SEC 3, 5, 6, and 7 increased comparatives to their SOC 2000 counterparts.

Looking again at table 2.15, RGSC for SOC 2000 is much more unevenly distributed than NS-SEC, with skilled manual workers making up 41.62 per cent of individuals and professionals only making up 4.30 per cent of individuals. This uneven distribution, on top of their analytical differences, presents some evidence to suggest that substantive findings of a sensitivity analysis could potentially find diverging findings. Compared to the SOC 2000 construction of RGSC, the SOC 90 construction also has some substantive deviations. RGSC 2 has a large decrease from 20.45 per cent to 7.74 per cent, RGSC 3M also has a large decrease from 41.62 per cent to 30.92 per cent. On the other hand, RGSC 4 has a large increase from 14.33 per cent to 34.95 per cent from SOC 2000 to SOC 90 construction.

Finally, CAMSIS for SOC 2000 construction has a mean of 44.57 and a standard deviation of 13.63 compared to CAMSIS for SOC 90 construction that has a mean of 42.04 and a standard deviation of 12.84.

Table 2.14 Descriptive Statistics for NCDS cohort Model

|  |  |  |
| --- | --- | --- |
| Table 2.13: Descriptive Statistics for Economic Activity | | |
|  | n | % |
| Continue Schooling or not after September when individuals are 16 |  |  |
| Don't Continue Schooling | 5,116 | 60.83% |
| Continue Schooling | 3,295 | 39.17% |
| Educational Attainment O-levels |  |  |
| <5 O-Levels | 5,426 | 64.51% |
| >5 O-Levels | 2,985 | 35.49% |
| Sex of Respondent |  |  |
| Female | 4,215 | 50.11% |
| Male | 4,196 | 49.89% |
| Housing Tenure of Respondent when Child |  |  |
| Own Home | 4,045 | 48.09% |
| Don't Own Home | 4,366 | 51.91% |
| NS-SEC Social Class of Father when Respondent Child SOC2000 |  |  |
| Large Employers and higher managerial occupations | 261 | 3.10% |
| Higher professional occupations | 410 | 4.87% |
| Lower Managerial and professional occupations | 1,038 | 12.34% |
| Intermediate occupations | 805 | 9.57% |
| Small employers and own account workers | 1,024 | 12.17% |
| Lower supervisory and technical occupations | 1,372 | 16.31% |
| Semi-routine occupations | 1,485 | 17.66% |
| Routine occupations | 2,016 | 23.97% |
| RGSC Social Class of Father when Respondent Child SOC2000 |  |  |
| Professional | 362 | 4.30% |
| Managerial and Technical | 1,720 | 20.45% |
| Skilled non-manual | 905 | 10.76% |
| Skilled manual | 3,501 | 41.62% |
| Partly skilled | 1,205 | 14.33% |
| Unskilled | 718 | 8.54% |
| NS-SEC Social Class of Father when Respondent Child SOC90 |  |  |
| Large Employers and higher managerial occupations | 9 | 0.11% |
| Higher professional occupations | 346 | 4.11% |
| Lower Managerial and professional occupations | 689 | 8.19% |
| Intermediate occupations | 870 | 10.34% |
| Small employers and own account workers | 678 | 8.06% |
| Lower supervisory and technical occupations | 1,414 | 16.81% |
| Semi-routine occupations | 2,060 | 24.49% |
| Routine occupations | 2,345 | 27.88% |
| RGSC Social Class of Father when Respondent Child SOC90 |  |  |
| Professional | 304 | 3.61% |
| Managerial and Technical | 651 | 7.74% |
| Skilled non-manual | 1,129 | 13.42% |
| Skilled manual | 2,601 | 30.92% |
| Partly skilled | 2,940 | 34.95% |
| Unskilled | 786 | 9.34% |
|  |  |  |
|  | Mean | SD |
| CAMSIS Score of Father when Respondent Child SOC2000 | 44.57 | 13.63 |
| CAMSIS Score of Father when Respondent Child SOC90 | 42.04 | 12.84 |
|  |  |  |
| n |  | 8411 |
| Data Source: NCDS [Sweeps 0-4] | | |

Table 2.14 provides descriptive statistics by the dependent variable of economic activity. An individual’s educational attainment is widely different when stratified by their economic activity. Those that don't continue schooling have a split of 89.68 per cent having achieved less than five O’levels compared to 10.32 per cent of their peers that achieved five or more O'levels. The reverse is true for those who continued schooling, whereby 74.57 per cent of individuals achieved five or more levels compared to 25.43 per cent that received less than five O'levels.

From observing the descriptive statistics, first transition is not stratified heavily by sex. Around 54.69 per cent of females and 45.31 per cent of males continue schooling. The lack of stratification could be the result of grouping multiple divergent transitionary pathways into a 'don't continue schooling' category.

Those who lived with parents who did not own their own homes make up the majority (63.84 per cent) of individuals who don't continue schooling compared to 36.16 per cent that do continue schooling. The reverse being true for those that own their own home, with 66.62 per cent continuing schooling.

Looking at NS-SEC for SOC 2000 construction, a majority of the sample located within the don't continue schooling category are situated within NS-SEC 7 and 6. Those that continue schooling see a larger makeup of children from social origins of NS-SEC 1.1,1.2, and 2. The same is true for the SOC 90 construction of NS-SEC. Though the SOC 90 construction does see a larger concentration of NS-SEC 6 and 7 not continuing schooling. Both RGSC measures demonstrate a strict manual/non-manual distinction whereby most individuals not continuing schooling come from RGSC 4 and 5. The RGSC SOC 2000 measure has a illustrates a larger manual/non-manual divide compared to its SOC 90 counterpart. CAMSIS measures demonstrates a higher overall mean for individuals that continue schooling compared to those that didn’t – however the SOC 2000 construction of CAMSIS has a much wider gap in mean between the two categories – a 10-point difference compared to the SOC 90 measures 8 points.

Table 2.15 Descriptive Statistics for NCDS model by Dependent Variable

|  |  |  |  |
| --- | --- | --- | --- |
| Descriptive Statistics by Economic Activity | | | |
|  | Continue Schooling or not after September when individuals are 16 | | |
|  | Don't Continue Schooling | Continue Schooling | Total |
| N | 5116 (60.83%) | 3295 (39.17%) | 8411 (100.00%) |
| Educational Attainment O-levels |  |  |  |
| <5 O-Levels | 4588 (89.68%) | 838 (25.43%) | 5426 (64.51%) |
| >5 O-Levels | 528 (10.32%) | 2457 (74.57%) | 2985 (35.49%) |
| Sex of Respondent |  |  |  |
| Female | 2413 (47.17%) | 1802 (54.69%) | 4215 (50.11%) |
| Male | 2703 (52.83%) | 1493 (45.31%) | 4196 (49.89%) |
| Housing Tenure of Respondent when Child |  |  |  |
| Own Home | 1850 (36.16%) | 2195 (66.62%) | 4045 (48.09%) |
| Don't Own Home | 3266 (63.84%) | 1100 (33.38%) | 4366 (51.91%) |
| NS-SEC Social Class of Father when Respondent Child SOC2000 |  |  |  |
| Large Employers and higher managerial occupations | 82 (1.60%) | 179 (5.43%) | 261 (3.10%) |
| Higher professional occupations | 82 (1.60%) | 328 (9.95%) | 410 (4.87%) |
| Lower Managerial and professional occupations | 363 (7.10%) | 675 (20.49%) | 1038 (12.34%) |
| Intermediate occupations | 358 (7.00%) | 447 (13.57%) | 805 (9.57%) |
| Small employers and own account workers | 671 (13.12%) | 353 (10.71%) | 1024 (12.17%) |
| Lower supervisory and technical occupations | 892 (17.44%) | 480 (14.57%) | 1372 (16.31%) |
| Semi-routine occupations | 1083 (21.17%) | 402 (12.20%) | 1485 (17.66%) |
| Routine occupations | 1585 (30.98%) | 431 (13.08%) | 2016 (23.97%) |
| RGSC Social Class of Father when Respondent Child SOC2000 |  |  |  |
| Professional | 72 (1.41%) | 290 (8.80%) | 362 (4.30%) |
| Managerial and Technical | 685 (13.39%) | 1035 (31.41%) | 1720 (20.45%) |
| Skilled non-manual | 416 (8.13%) | 489 (14.84%) | 905 (10.76%) |
| Skilled manual | 2457 (48.03%) | 1044 (31.68%) | 3501 (41.62%) |
| Partly skilled | 898 (17.55%) | 307 (9.32%) | 1205 (14.33%) |
| Unskilled | 588 (11.49%) | 130 (3.95%) | 718 (8.54%) |
| NS-SEC Social Class of Father when Respondent Child SOC90 |  |  |  |
| Large Employers and higher managerial occupations | 3 (0.06%) | 6 (0.18%) | 9 (0.11%) |
| Higher professional occupations | 79 (1.54%) | 267 (8.10%) | 346 (4.11%) |
| Lower Managerial and professional occupations | 221 (4.32%) | 468 (14.20%) | 689 (8.19%) |
| Intermediate occupations | 332 (6.49%) | 538 (16.33%) | 870 (10.34%) |
| Small employers and own account workers | 438 (8.56%) | 240 (7.28%) | 678 (8.06%) |
| Lower supervisory and technical occupations | 890 (17.40%) | 524 (15.90%) | 1414 (16.81%) |
| Semi-routine occupations | 1355 (26.49%) | 705 (21.40%) | 2060 (24.49%) |
| Routine occupations | 1798 (35.14%) | 547 (16.60%) | 2345 (27.88%) |
| RGSC Social Class of Father when Respondent Child SOC90 |  |  |  |
| Professional | 67 (1.31%) | 237 (7.19%) | 304 (3.61%) |
| Managerial and Technical | 191 (3.73%) | 460 (13.96%) | 651 (7.74%) |
| Skilled non-manual | 476 (9.30%) | 653 (19.82%) | 1129 (13.42%) |
| Skilled manual | 1910 (37.33%) | 691 (20.97%) | 2601 (30.92%) |
| Partly skilled | 1892 (36.98%) | 1048 (31.81%) | 2940 (34.95%) |
| Unskilled | 580 (11.34%) | 206 (6.25%) | 786 (9.34%) |
| CAMSIS Score of Father when Respondent Child SOC2000 | 40.49 (11.27) | 50.90 (14.53) | 44.57 (13.63) |
| CAMSIS Score of Father when Respondent Child SOC90 | 38.93 (10.53) | 46.87 (14.50) | 42.04 (12.84) |
| Data Source: NCDS [Sweeps 0-4] | | | |

Looking in further detail on the analytical construction of each of the three social stratification variables a cross tabulation is created for both NS-SEC and RGSC measures – and summary statistics provided for CAMSIS. This comparison of measures illustrates the trends and patterns that are associated with creating a social stratification measure using two distinct SOC codes. Table 2.15 details a cross-tabulation of NS-SEC by SOC construction. Looking at the diagonals demonstrates how many observations share the same NS-SEC category for both the SOC 2000 and SOC 90 constructions. No diagonal has less than 60 per cent overlap. The lowest overlap occurs in NS-SEC 3 and NS-SEC 6 with 60.29 per cent and 60.78 per cent respectively. The largest single overlap occurs at NS-SEC 1.1 with 100 per cent overlap.

Table 2.16 Descriptive Statistics Crosstab of NS-SEC for NCDS model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Descriptive Statistics comparing NS-SEC by SOC2000 and SOC90 codes | | | | | | | | | |
|  | NS-SEC Social Class of Father when Respondent Child SOC90 | | | | | | | | |
|  | Large Employers and higher managerial occupations | Higher professional occupations | Lower Managerial and professional occupations | Intermediate occupations | Small employers and own account workers | Lower supervisory and technical occupations | Semi-routine occupations | Routine occupations | Total |
| N | 9 (0.11%) | 346 (4.11%) | 689 (8.19%) | 870 (10.34%) | 678 (8.06%) | 1414 (16.81%) | 2060 (24.49%) | 2345 (27.88%) | 8411 (100.00%) |
| NS-SEC Social Class of Father when Respondent Child SOC2000 |  |  |  |  |  |  |  |  |  |
| Large Employers and higher managerial occupations | 9 (100.00%) | 18 (5.20%) | 19 (2.76%) | 87 (10.00%) | 0 (0.00%) | 7 (0.50%) | 121 (5.87%) | 0 (0.00%) | 261 (3.10%) |
| Higher professional occupations | 0 (0.00%) | 285 (82.37%) | 78 (11.32%) | 46 (5.29%) | 1 (0.15%) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) | 410 (4.87%) |
| Lower Managerial and professional occupations | 0 (0.00%) | 43 (12.43%) | 526 (76.34%) | 184 (21.15%) | 19 (2.80%) | 8 (0.57%) | 174 (8.45%) | 84 (3.58%) | 1038 (12.34%) |
| Intermediate occupations | 0 (0.00%) | 0 (0.00%) | 13 (1.89%) | 528 (60.69%) | 61 (9.00%) | 86 (6.08%) | 103 (5.00%) | 14 (0.60%) | 805 (9.57%) |
| Small employers and own account workers | 0 (0.00%) | 0 (0.00%) | 53 (7.69%) | 11 (1.26%) | 511 (75.37%) | 267 (18.88%) | 179 (8.69%) | 3 (0.13%) | 1024 (12.17%) |
| Lower supervisory and technical occupations | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) | 14 (1.61%) | 29 (4.28%) | 984 (69.59%) | 141 (6.84%) | 204 (8.70%) | 1372 (16.31%) |
| Semi-routine occupations | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) | 45 (6.64%) | 1 (0.07%) | 1252 (60.78%) | 187 (7.97%) | 1485 (17.66%) |
| Routine occupations | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) | 12 (1.77%) | 61 (4.31%) | 90 (4.37%) | 1853 (79.02%) | 2016 (23.97%) |
| Data Source: NCDS [Sweeps 0-4] | | | | | | | | | |

Note: % at column. Pearson Chi2 p<0.00

Table 2.17 Descriptive Statistics Crosstab of RGSC for NCDS model

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Descriptive Statistics comparing RGSC by SOC2000 and SOC90 codes | | | | | | | |
|  | RGSC Social Class of Father when Respondent Child SOC90 | | | | | | |
|  | Professional | Managerial and Technical | Skilled non-manual | Skilled manual | Partly skilled | Unskilled | Total |
| N | 304 (3.61%) | 651 (7.74%) | 1129 (13.42%) | 2601 (30.92%) | 2940 (34.95%) | 786 (9.34%) | 8411 (100.00%) |
| RGSC Social Class of Father when Respondent Child SOC2000 |  |  |  |  |  |  |  |
| Professional | 268 (88.16%) | 73 (11.21%) | 20 (1.77%) | 0 (0.00%) | 1 (0.03%) | 0 (0.00%) | 362 (4.30%) |
| Managerial and Technical | 36 (11.84%) | 542 (83.26%) | 446 (39.50%) | 6 (0.23%) | 651 (22.14%) | 39 (4.96%) | 1720 (20.45%) |
| Skilled non-manual | 0 (0.00%) | 3 (0.46%) | 652 (57.75%) | 42 (1.61%) | 186 (6.33%) | 22 (2.80%) | 905 (10.76%) |
| Skilled manual | 0 (0.00%) | 32 (4.92%) | 10 (0.89%) | 2015 (77.47%) | 1349 (45.88%) | 95 (12.09%) | 3501 (41.62%) |
| Partly skilled | 0 (0.00%) | 1 (0.15%) | 1 (0.09%) | 191 (7.34%) | 753 (25.61%) | 259 (32.95%) | 1205 (14.33%) |
| Unskilled | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) | 347 (13.34%) | 0 (0.00%) | 371 (47.20%) | 718 (8.54%) |
| Data Source: NCDS [Sweeps 0-4] | | | | | | | |

Note: % at column. Pearson Chi2 p<0.00

For the RGSC schema in table 2.16, looking at the diagonals, whilst RGSC 1-3M have an overlap of over 50 per cent both RGSC 4 and 5 have overlaps of 25.61 per cent and 47.20 per cent respectively. This demonstrates a radical divergence in operationalisation of the lowest two RGSC categories from SOC 90 to SOC 2000 and there is an expectation of diverging substantive findings.

Table 2.18 Descriptive Statistics of CAMSIS for NCDS model

|  |  |
| --- | --- |
| CAMSIS2000 | |
| *Mean* | 44.57 |
| *Standard Deviation* | 13.63 |
| CAMSIS90 | |
| *Mean* | 42.04 |
| *Standard Deviation* | 12.84 |
| N | 8411 |
| Data Source: NCDS [Sweeps 0-4] | |

Note: Correlation matrix = 0.81

For CAMSIS for SOC 2000 construction, with a base total mean of 44.57 those that continued schooling had a mean of 50.90 and those that didn't continue schooling had a mean of 40.49. The SOC 90 construction of CAMSIS echoes this pattern but there is a smaller difference between the mean of continuing schooling and those that didn't continue schooling.

### Initial Model

The first set of analyses estimates a logistic regression model with NS-SEC as the chosen social stratification measure (RGSC and CAMSIS will be introduced in a sensitivity analysis later). Table 2.18 details the deviance, change in deviance, change in degrees of freedom, and McFadden’s Pseudo , AIC, and BIC measures to compare the null model with models of one explanatory variable. Table 2.19 details the exact statistics but through a sequential building of the null model with each subsequent independent variable added.

This model has been evaluated for the goodness of fit of two competing statistical models based on the ratio of their likelihoods in a likelihood-ratio test and again with a Wald test. Both found that the hypothesis that all the coefficients associated with educational attainment, sex, tenure, and NS-SEC are simultaneously equal to 0 can be rejected at the 0.01 level.

Table 2.19 Model Building Statistics for NCDS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d. f. (from Null) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 11262.73 | - | - | - | 11264.73 | 11271.77 |
| Null Model + Educational Attainment | 7455.83 | 3,806.90 | 2 | 0.34 | 7459.83 | 7473.91 |
| Null Model + Sex | 11217.31 | 45.42 | 2 | 0.00 | 11221.31 | 11235.38 |
| Null Model + Tenure | 10507.00 | 755.73 | 2 | 0.07 | 10511.00 | 10525.07 |
| Null Model + NS-SEC (SOC 2000) | 10107.27 | 1,155.46 | 8 | 0.10 | 10123.27 | 10179.57 |

Explanatory variables are entered sequentially in the subsequent logistic models following the (Gayle and Lambert, 2009) example.

Table 2.20 Sequential Model Building Statistics for NCDS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 11262.73 | - | - | - | 11264.73 | 11271.77 |
| Null Model + Educational Attainment | 7455.83 | 3,806.90 | 2 | 0.34 | 7459.83 | 7473.91 |
| Null Model + Educational Attainment + Sex | 7400.72 | 55.11 | 2 | 0.34 | 7406.73 | 7427.84 |
| Null Model + Educational Attainment + Sex + Tenure | 7192.56 | 208.16 | 2 | 0.36 | 7200.56 | 7228.71 |
| Null Model + Educational Attainment + Sex + Tenure + NS-SEC (SOC 2000) | 6987.62 | 204.94 | 8 | 0.38 | 7009.62 | 7087.03 |

The model fit statistics demonstrate that there are typically distributed residuals and that the model is correctly specified. Table 2.20 suggests that deviance is reduced by 4,275.11 from the null for the full proposed model. AIC and BIC statistics also suggest that the full model best fits those entered. Finally, the full model presents an of 0.38. In other words, the full model explains 38 per cent of the variance of first transitions, leaving 52 per cent unexplained. The following analysis with the full model is a complete records analysis with 8,411 observations.

Following from a discussion on rescaling effects in non-linear models, table 2.20 details the KHB method and decomposes the direct and indirect effects for each nested model. KHB only works with at least one independent variable as a point of comparison. Therefore model 1 will be a null model + educational attainment + sex, with each model adding a variable until reaching model 3 which will be all analytical variables.

Table 2.21 KHB Nested Regression Comparisons

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Model 1 | | Model 2 | | Model 3 | |
|  |  | **Log Odds** | **SE** | **Log Odds** | **SE** | **Log Odds** | **SE** |
| Educational Attainment | *Reduced* | 3.27\*\*\* | (0.06) | 3.38\*\*\* | (0.07) | 3.25\*\*\* | (0.07) |
| *Full* | 3.26\*\*\* | (0.06) | 3.12\*\*\* | (0.06) | 2.99\*\*\* | (0.06) |
| *Difference* | 0.01 | (0.01) | 0.26\*\*\* | (0.02) | 0.26\*\*\* | (0.03) |
| Sex | *Reduced* |  |  | -0.46\*\*\* | (0.06) | -0.50\*\*\* | (0.06) |
| *Full* |  |  | -0.48\*\*\* | (0.06) | -0.50\*\*\* | (0.06) |
| *Difference* |  |  | 0.03\* | (0.01) | 0.00 | (0.02) |
| Housing Tenure | *Reduced* |  |  |  |  | -0.88\*\*\* | (0.06) |
| *Full* |  |  |  |  | -0.63\*\*\* | (0.06) |
| *Difference* |  |  |  |  | -0.25\*\*\* | (0.03) |

Table 2.22 KHB Summary Statistics of Nested Regression Comparisons

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Model 1 | | | Model 2 | | | Model 3 | | |
|  | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** |
| Educational Attainment | 1.00 | 0.27 | 1.01 | 1.08 | 7.57 | 1.04 | 1.09 | 8.03 | 1.04 |
| Sex |  |  |  | 0.94 | -5.94 | 1.03 | 0.99 | -0.78 | 1.04 |
| Housing Tenure |  |  |  |  |  |  | 1.40 | 28.38 | 1.01 |

Tables 2.20 and 2.21 provide detailed decompositions of the confounding and rescaling factors of comparing nested models using the same analytical sample. We can see from table 2.20 model 1 that the log odds of continuing schooling are increased by 3.27 for those with five or more O’levels compared to those with less than five O’levels. Controlling for sex, the direct effect of educational attainment is reduced to 3.26 leaving an indirect effect of 0.01. Using table 2.21 this can be interpreted precisely. The total effect of educational attainment is 1.00 times larger than the direct effect (no difference) when controlling for sex. The confounding percentage for model 1 indicates that 0.27 percent of the total effects of receiving five or more O’levels is due to the additional explanatory variables added to the model. The total effect of educational attainment is 1.01 times larger than the direct effect when controlling for sex.

Moving to model 3 with the addition of NS-SEC to the analytical model, the total effect of educational attainment is 1.09 times larger than the direct effect when controlling for all other variables. The confounding percentage indicates that 8.03 percent of the total effects of receiving five or more O’levels is due to the additional explanatory variable added to the model. The total effect of educational attainment is 1.04 times larger than the direct effect when controlling for all other variables. No single variable within model 3 has a substantively large rescaling factor, meaning that the change in total effect can be almost entirely attributed to the mediating effect of additional variables. Housing tenure in model 3 has a large confounding percentage at 28.38 percent. This confounding percentage indicates that 28.38 percent of the total effects of not owning your own home is due to the additional explanatory variables added to the model. That is a substantive amount of total effect size change due to the mediating effect of NS-SEC.

The results of the logistic regression model are reported in Table 2.22. Log odds, average marginal effects, and quasi-variance statistics are reported. It is impossible to ascertain the significance of variables' parameters other than the reference category (Firth, 2003). This is known as the reference category problem (Connelly, Gayle and Lambert, 2016). Quasi-variances were considered to overcome this reference category problem; these are detailed for NS-SEC (the only variable that can provide quasi-variance statistics) in Table 2.23. Quasi-variances are reported alongside the standard outputs of log odds and average marginal effects to overcome the reference category problem. Quasi-variances are normally produced within Stata using the custom Stata command ‘qv’ (Chen, 2014)[[17]](#footnote-17).

Table 2.23 Analytical Model for NCDS

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | NS-SEC | | | Average Marginal Effects | | Quasi-variance | | |
| Economic Activity: ‘Don’t Continue Schooling’ Reference Category | **Log Odds** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **S.E.** | **LB** | **UB** |
| Continue Schooling |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. |  |  |  |  |  |  |  |
| *Five or More O’levels* | 2.98 | (0.07) | \*\*\* | 0.56 | (0.01) |  |  |  |
| Sex |  |  |  |  |  |  |  |  |
| *Female* | Ref. |  |  |  |  |  |  |  |
| *Male* | -0.50 | (0.06) | \*\*\* | -0.06 | (0.01) |  |  |  |
| Housing Tenure |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. |  |  |  |  |  |  |  |
| *Do not Own Home* | -0.63 | (0.06) | \*\*\* | -0.08 | (0.01) |  |  |  |
| NS-SEC (SOC 2000) |  |  |  |  |  |  |  |  |
| *1.1 Large employers and higher managerial occupations* | 0.02 | (0.19) |  | 0.00 | (0.03) | 0.17 | -0.36 | 0.40 |
| *1.2 higher professional occupations* | 0.48 | (0.17) | \*\*\* | 0.08 | (0.03) | 0.15 | 0.14 | 0.82 |
| *2 Lower managerial and professional occupations* | Ref. |  |  |  |  | 0.09 | -0.19 | 0.19 |
| *3 Intermediate occupations* | -0.25 | (0.13) | \* | -0.04 | (0.02) | 0.09 | -0.46 | -0.04 |
| *4 Small employers and own account workers* | -0.89 | (0.12) | \*\*\* | -0.13 | (0.02) | 0.09 | -1.08 | -0.70 |
| *5 Lower supervisory and technical occupations* | -0.76 | (0.11) | \*\*\* | -0.11 | (0.02) | 0.07 | -0.93 | -0.60 |
| *6 Semi-routine occupations* | -0.89 | (0.11) | \*\*\* | -0.13 | (0.02) | 0.07 | -1.05 | -0.72 |
| *7 Routine occupations* | -1.11 | (0.11) | \*\*\* | -0.16 | (0.02) | 0.07 | -1.26 | -0.95 |
| Intercept | -0.40 | (0.09) | \*\*\* | (.) | (.) | (.) | (.) | (.) |
| Number of observations | 8411 | | | | | | | |
| McFadden’s | 0.38 | | | | | | | |
| McFadden’s Adjusted Pseudo | 0.38 | | | | | | | |
| Cox-Snell Pseudo | 0.40 | | | | | | | |
| Nagelkerke Pseudo | 0.53 | | | | | | | |
| Tjur’s | 0.46 | | | | | | | |
| AIC | 7009.62 | | | | | | | |
| BIC | 7087.03 | | | | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: NCDS [Sweeps 0-4]  Note: Complete Records Analysis | | | | | | | | |

The output in table 2.21 demonstrates that for those individuals who received five or more O'levels compared to those that received less than five O’levels had an increased log odd of continuing schooling. Translated into average marginal effects, this represents an increased probability of 56 percentage points to continue schooling if individuals received five or more O'levels. This model demonstrates that educational attainment is by far the largest single influence on continuing education for the NCDS cohort. The size of the effect suggests that educational attainment appears to be a major roadblock in a young persons first major transition. Not attaining five or more O'levels appears to have a major decreased likelihood of continuing schooling. This provides evidence for a pre-credentialled labour market – young people were not continuing schooling en masse, there were some requirements to fulfil in order to do so, prior educational attainment being one such requirement.

Sex is found to be statistically significant. Men in the NCDS have a decreased log odds of continuing schooling versus women. Translated into average marginal effects this is a 6 percentage points decrease of continuing schooling. This sexed divide may be a consequence of traditional gender roles that influenced men entering employment as soon as possible and for women to become homemakers and wives, thus being able to continue schooling whilst men had more restrictive expectations placed upon them in terms of economic activity.

Moving on to housing tenure, individuals that grew up in homes not owned by their parents have a decreased odds of continuing schooling compared to individuals that grew up in homes owned by their parents. Translated to average marginal effects this represents an 8 percentage points decrease. This housing tenure effect could be a case of mediating a social class effect, as witnessed to some extent with the KHB decomposition. However this is not the entire total effect of housing tenure meaning that in of itself housing tenure is an influence upon youths first major transition. Housing tenure could be capturing a more nuanced neighbourhood effect – individuals growing up in neighbourhoods that all own their own homes versus on a council estate etc.

Finally moving on to NS-SEC, classes 1.2, and 3-7 are statistically significant and will be interpreted substantively. Whilst individuals from social origins NS-SEC 1.2 compared to NS-SEC 2 had an increased log odds or 8 percentage point increase of continuing schooling individuals in NS-SEC 3-7 all had decreased log odds in a near monotonic pattern (the exception being NS-SEC 4) from 4 to 16 percentage point decrease of continuing schooling compared to those from social origins NS-SEC 2. A full graphical breakdown of the coefficients for all variables can be found in figure 2.6. All but NS-SEC 4 follow a monotonic decrease in likelihood to continue schooling. This is suggestive of a strong class based effect on continuing schooling. The non-monotonic ‘bump’ witnessed in NS-SEC 4 can be explained by the makeup of that particular category. NS-SEC 4 is made up of small employers and own account holders –- a set of occupations that do not rely upon traditional education structures for success. NS-SEC 4 being the single largest group in NS-SEC to have a decreased likelihood of continuing schooling is understandable in this context.

A graph with lines and dots

Description automatically generated

Figure ‎2.7 Coefficient plot of analytical model

There are small divergences across various measures. All measures suggest a range of variance explained from 38-53 per cent. All agree that this model explains a large amount of variance.

To understand this in a more manageable format, graphs are produced to aid in substantive interpretation. First a graph presenting log odds and quasi-variance statistics for NS-SEC are provided. Then NS-SEC is graphed using both predictive probabilities and average marginal effects. Finally, all other variables are graphed using predictive probabilities. Graphing predicted probabilities by a variable rather than looking at a table with variables grouped by outcome variable allows for each variable to have cross-outcome group trends to be compared.

Reflecting on figure 2.7, the underlying trend for the quasi-variance compared to the log odds counterparts is that coefficients remain constant whilst standard errors and confidence intervals are slightly reduced – this is a direct result of resolving the reference category problem. There are very minor changes comparing log odds to quasi-variance statistics. The largest reduction in standard errors comparative to log odds is concentrated in NS-SEC 4-7.

A graph with red and black lines

Description automatically generated

Figure ‎2.8 Log odds versus Quasi-Variance Statistics for NCDS model (NS-SEC)

Focusing on NS-SEC, figure 2.8 depicts the predicted probabilities at means of economic activity alongside the average marginal effects of NS-SEC compared to the reference category of NS-SEC 2. Both graphs are represented using the same common y axis to aid interpretation. With respect to predicted probabilities except for 1.1-1.2, where there is a slight increase in people continuing schooling, there is a near general monotonic decreased trend for individuals to continue schooling from NS-SEC 1.2-7. Moving on to the average marginal effects of NS-SEC there is a monotonic decrease from NS-SEC 1.1 to NS-SEC 4, whereby from NS-SEC 4 to 6 there is a flatlining followed by a small decrease from NS-SEC 6 to NS-SEC 7. The largest average increased marginal probability reported in this graph relates to an 8 per cent increased probability of continuing schooling for NS-SEC 1.2 compared to NS-SEC 2. The largest average decreased marginal probability reported in the graph relates to a 13 per cent decreased probability of continuing schooling for NS-SEC 7 compared to NS-SEC 2.

A screenshot of a computer screen

Description automatically generated

Figure ‎2.9 Predictive and AMEs of NS-SEC for NCDS Model

For all other variables in the model, they are solely graphically visualised through predictive probabilities - average marginal effects are reported at table 2.20. Figures 2.9-2.11 present the predicted probabilities of each analytical variable not spoken about thus far. Starting with figure 2.10, the predicted probabilities of educational attainment are graphed. Young people in the NCDS who received less than five O’levels have a drastically lower predicted probability of continuing schooling compared to those young people that received five or more O’levels.

A white rectangular object with black text

Description automatically generated

Figure ‎2.10 Predictive Margins of Educational Attainment for NCDS model

Figure 2.11 documents the predictive probabilities of sex. Whilst not as large as educational attainment, there is still a clear gap in the predicted probabilities of men and women continuing education. Women have a higher predicted probability of continuing schooling than men.

A white background with black text

Description automatically generated

Figure ‎2.11 Predictive Margins of Sex for NCDS model

Figure 2.12 documents the predictive margins of housing tenure in the NCDS. Young people that grew up in homes owned by their parents have a larger predicted probability of continuing schooling compared to young people that grew up in homes that were not owned by their parents.

A white paper with black text

Description automatically generated

Figure ‎2.12 Predictive Margins of Housing Tenure for NCDS model

### Sensitivity Analysis of Social Stratification Measures using NCDS

There are a variety of social stratification measures used by social scientists. The choice of which social stratification measure to use within a model is dependent upon researcher discretion. Some researchers may have theoretical justifications for using a particular social stratification measure over others, other researchers may simply use the variables provided to them within the given dataset they are using. When referring to matters of social inequality or social class, the particularities of what social stratification measure a particular model uses is often times not made clear. It is difficult to assess if using one social stratification measure over another in a particular model would have a substantively different interpretation. This is why sensitivity analyses are important in social research. This section employs a sensitivity analyses of social stratification measures employed within this analysis – comparing three social stratification measures for their substantive interpretation to understand and recognise and differences. These differences will then be discussed.

Sensitivity analysis is not common practice within social stratification research (Lambert and Barnett, 2021). A sensitivity analysis of social stratification measures provides the most well-informed assessment about the role different social stratification measures have on the substantive interpretation of a given model. NS-SEC, CAMSIS, and RGSC are three of these measures. The analytical distinctions between these three measures have already been discussed. Given the historical nature of the NCDS cohort, a sensitivity analysis would provide an exciting insight into the temporal sensitivity of these socio-economic measures, as well as presenting results that demonstrate the best model fit. The subsequent sensitivity analysis will compare like-for-like models of economic activity, each using a different socio-economic measure. The base model – NS-SEC – is then compared to the CAMSIS and RGSC models. While it is not appropriate to compare log odds across regression models (Connelly, Gayle and Lambert, 2016). Employment of KHB decomposition provides some background on the ability to compare models following substantive conclusions. Goodness-of-fit statistics are provided and are assessed via AIC, BIC, and a range of measures.

There are strong correlations between parental social class measures. Parental NS-SEC and Parental RGSC have a significant Chi-Square statistic at the p<0.01 level. Parental NS-SEC and CAMSIS have a significant Anova at the p<0.01 level.

Three separate multinomial logistic regressions are presented in Table 2.24. The first model has been described at length in the previous section and uses NS-SEC. The second model uses CAMSIS, and the third uses RGSC. These models are all presented using log odds and average marginal effects to enhance interpretation and comparison.

As with the base NS-SEC model above, model building statistics for both RGSC and CAMSIS measures are provided in tables 2.23-2.26.

Table 2.24 Model building statistics of RGSC for NCDS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d. f. (from Null) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 11262.73 | - | - | - | 11264.73 | 11271.77 |
| Null Model + Educational Attainment | 7455.83 | 3,806.90 | 2 | 0.34 | 7459.83 | 7473.91 |
| Null Model + Sex | 11217.31 | 45.42 | 2 | 0.00 | 11221.31 | 11235.38 |
| Null Model + Tenure | 10507.00 | 755.73 | 2 | 0.07 | 10511.00 | 10525.07 |
| Null Model + RGSC (SOC 2000) | 10236.07 | 1,026.66 | 6 | 0.09 | 10248.06 | 10290.29 |

Table 2.25 Model building statistics of CAMSIS for NCDS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d. f. (from Null) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 11262.73 | - | - | - | 11264.73 | 11271.77 |
| Null Model + Educational Attainment | 7455.83 | 3,806.90 | 2 | 0.34 | 7459.83 | 7473.91 |
| Null Model + Sex | 11217.31 | 45.42 | 2 | 0.00 | 11221.31 | 11235.38 |
| Null Model + Tenure | 10507.00 | 755.73 | 2 | 0.07 | 10511.00 | 10525.07 |
| Null Model + CAMSIS (SOC 2000) | 1,212.96 | 1,059.64 | 2 | 0.11 | 10053.77 | 10067.84 |

Table 2.26 Sequential Model building statistics of RGSC for NCDS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 11262.73 | - | - | - | 11264.73 | 11271.77 |
| Null Model + Educational Attainment | 7455.83 | 3,806.90 | 2 | 0.34 | 7459.83 | 7473.91 |
| Null Model + Educational Attainment + Sex | 7400.72 | 55.11 | 2 | 0.34 | 7406.73 | 7427.84 |
| Null Model + Educational Attainment + Sex + Tenure | 7192.56 | 208.16 | 2 | 0.36 | 7200.56 | 7228.71 |
| Null Model + Educational Attainment + Sex + Tenure + RGSC (SOC 2000) | 7024.88 | 167.68 | 6 | 0.38 | 7042.88 | 7106.21 |

Table 2.27 Sequential Model building statistics of CAMSIS for NCDS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 11262.73 | - | - | - | 11264.73 | 11271.77 |
| Null Model + Educational Attainment | 7455.83 | 3,806.90 | 2 | 0.34 | 7459.83 | 7473.91 |
| Null Model + Educational Attainment + Sex | 7400.72 | 55.11 | 2 | 0.34 | 7406.73 | 7427.84 |
| Null Model + Educational Attainment + Sex + Tenure | 7192.56 | 208.16 | 2 | 0.36 | 7200.56 | 7228.71 |
| Null Model + Educational Attainment + Sex + Tenure + CAMSIS (SOC 2000) | 6984.95 | 207.61 | 2 | 0.38 | 6994.95 | 7030.13 |

KHB analysis is employed to check how much of the changing total effect size of log odds within each analytical model is representative of rescaling over mediating of additional variables. Tables 2.27-28 details a similar phenomenon for each full analytical models proposed for a sensitivity analysis – model 1 uses NS-SEC, model 2 uses RGSC, and model 3 uses CAMSIS. All three models use the same analytical sample. The table below provides detail on any variation in the levels of rescaling versus mediation for each social stratification variable added to the proposed model. These differences will be briefly discussed prior to substantive interpretation of each analytical model.

Whilst there are minor differences in the difference in total effect sizes across each model in table 2.27, the resulting summary statistics produced in table 2.28 confirm that these differences are indeed small enough to lack a substantive difference between models. All three models have very small total effect changes that relate to rescaling, the majority of total effect size differences in each model comes from the mediating aspect of additional variables within their separate models.

Table 2.28 KHB method of Nested Regression Models for Social Stratification Analysis

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Model 1 | | Model 2 | | Model 3 | |
|  |  | **Log Odds** | **SE** | **Log Odds** | **SE** | **Log Odds** | **SE** |
| Educational Attainment | *Reduced* | 3.25\*\*\* | (0.07) | 3.22\*\*\* | (0.07) | 3.25\*\*\* | (0.07) |
| *Full* | 2.99\*\*\* | (0.06) | 3.00\*\*\* | (0.06) | 2.97\*\*\* | (0.06) |
| *Difference* | 0.26\*\*\* | (0.03) | 0.22\*\*\* | (0.02) | 0.27\*\*\* | (0.03) |
| Sex | *Reduced* | -0.50\*\*\* | (0.06) | -0.50\*\*\* | (0.06) | -0.50\*\*\* | (0.06) |
| *Full* | -0.50\*\*\* | (0.06) | -0.50\*\*\* | (0.06) | -0.51\*\*\* | (0.06) |
| *Difference* | 0.00 | (0.02) | 0.01 | (0.02) | 0.01 | (0.02) |
| Housing Tenure | *Reduced* | -0.88\*\*\* | (0.06) | -0.88\*\*\* | (0.06) | -0.88\*\*\* | (0.06) |
| *Full* | -0.63\*\*\* | (0.06) | -0.64\*\*\* | (0.06) | -0.60\*\*\* | (0.06) |
| *Difference* | -0.25\*\*\* | (0.03) | -0.24\*\*\* | (0.03) | -0.29\*\*\* | (0.03) |

Table 2.29 KHB method Summary Statistics for Nested Regression Models for Social Stratification Analysis

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Model 1 | | | Model 2 | | | Model 3 | | |
|  | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** |
| Educational Attainment | 1.09 | 8.03 | 1.04 | 1.07 | 6.78 | 1.03 | 1.09 | 8.45 | 1.04 |
| Sex | 0.99 | -0.78 | 1.04 | 0.99 | -1.35 | 1.03 | 0.97 | -2.75 | 1.03 |
| Housing Tenure | 1.40 | 28.38 | 1.01 | 1.37 | 27.06 | 1.01 | 1.48 | 32.45 | 1.01 |

Moving on to interpretation of the sensitivity analysis of social stratification measures, there are key substantive differences in the three models. Model one, or the NS-SEC model will not be substantively interoperated here as it has already been extensively interpretated in table 2.21 and its subsequent discussion. As such, this model will only be discussed as far as it contrasts and compares with the CAMSIS and RGSC models. Unsurprisingly, all non-social stratification-based measures provide near-identical substantive findings: educational attainment, sex, and housing tenure.

A comparison of the NS-SEC and RGSC models identifies a remarkable substantive similarity. Just as NS-SEC follows a decreased near monotonic pattern from NS-SEC 1.2 to 7, the RGSC model is statistically significant at RGSC 1, 3M, and 4. These also follow a decreased monotonic pattern. Those individuals from social origins RGSC 1 compared to RGSC 2 had an increased likelihood of continuing schooling. Translated to average marginal effects this represents a 11 percentage point increase of continuing schooling over not continuing schooling. Those individuals from social origins RGSC 3M and 4 had a decreased likelihood of continuing schooling compared to their RGSC 2 peers. This translates to a decreased 9 percentage point and 11 percentage point of continuing schooling versus not continuing schooling. The NS-SEC and RGSC models both present data that is substantively similar. The CAMSIS model however diverges from both. Whilst CAMSIS is statistically significant and represents an increased log odds of continuing schooling for each point increase of CAMSIS when translated to average marginal effects this represents a 0 percentage point increase. It is not overly surprising that NS-SEC and RGSC present similar substantive results whilst CAMSIS does not considering the first two social stratification measures are measures of social class and CAMSIS is not, it is in fact a measure of social distance.

All three models have near identical measures. Both the McFadden's, McFadden's Adjusted, and Cox-Snell Pseudo all report around 38 per cent of variance explained. The Nagelkerke Pseudo has an increased measure of 53 per cent variance explained and Tjur's reports 45 per cent variance explained. All measures agree that this model explains a large amount of variance.

The goodness-of-fit statistics are similar for all three models. Differences in measures exist, but the minor nature of these differences indicates that the amount of variance explained across the three models remains consistent. AIC and BIC differences are also minor. The most parsimonious model is the CAMSIS model when using AIC and BIC. Considering that BIC penalises models for estimating additional parameters, it is not entirely surprising that it considers the CAMSIS a better fit than the NS-SEC or RGSC schema. These differences are, however, minimal. Whilst the goodness-of-fit statistics presented are interesting, the primary purpose of this sensitivity analysis was not to find the most parsimonious model, it was in fact to understand, if any, the substantive distinctions between social stratification measures. As such, going forward, the preferred model of choice for subsequent analysis will be the NS-SEC model.

Table 2.30 Sensitivity analysis of social stratification measures for NCDS model

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | NS-SEC | | | Average Marginal Effects | | Quasi-variance | | | RGSC | | | Average Marginal Effects | | Quasi-variance | | | CAMSIS | | | Average Marginal Effects | |
| Economic Activity: ‘Don’t Continue Schooling’ Reference Category | **Log odds** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **S.E.** | **LB** | **UB** | **Log odds** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **S.E.** | **LB** | **UB** | **Log odds** | **S.E.** | **Sig.** | **Prob.** | **S.E.** |
| Continue Schooling |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Five or More O’levels* | 2.98 | (0.07) | \*\*\* | 0.59 | (0.01) | (.) | (.) | (.) | 3.00 | (0.06) | \*\*\* | 0.60 | (0.01) | (.) | (.) | (.) | 2.97 | (0.06) | \*\*\* | 0.59 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Male* | -0.50 | (0.06) | \*\*\* | -0.06 | (0.01) | (.) | (.) | (.) | -0.50 | (0.06) | \*\*\* | -0.06 | (0.01) | (.) | (.) | (.) | -0.51 | (0.06) | \*\*\* | -0.07 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Do not Own Home* | -0.63 | (0.06) | \*\*\* | -0.08 | (0.01) | (.) | (.) | (.) | -0.64 | (0.06) | \*\*\* | -0.09 | (0.01) | (.) | (.) | (.) | -0.59 | (0.06) | \*\*\* | -0.08 | (0.01) |
| NS-SEC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1.1 Large employers and higher managerial occupations* | 0.02 | (0.19) |  | 0.00 | (0.03) | 0.17 | -0.36 | 0.40 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1.2 higher professional occupations* | 0.48 | (0.17) | \*\*\* | 0.08 | (0.03) | 0.15 | 0.14 | 0.82 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *2 Lower managerial and professional occupations* | Ref. |  |  |  |  | 0.09 | -0.19 | 0.19 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *3 Intermediate occupations* | -0.25 | (0.13) | \* | -0.04 | (0.02) | 0.09 | -0.46 | -0.04 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *4 Small employers and own account workers* | -0.89 | (0.12) | \*\*\* | -0.13 | (0.02) | 0.09 | -1.08 | -0.70 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *5 Lower supervisory and technical occupations* | -0.76 | (0.11) | \*\*\* | -0.11 | (0.02) | 0.07 | -0.93 | -0.60 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *6 Semi-routine occupations* | -0.89 | (0.11) | \*\*\* | -0.13 | (0.02) | 0.07 | -1.05 | -0.72 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *7 Routine occupations* | -1.11 | (0.11) | \*\*\* | -0.16 | (0.02) | 0.07 | -1.26 | -0.95 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| RGSC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1 Professional occupations* |  |  |  |  |  |  |  |  | 0.68 | (0.18) | \*\*\* | 0.11 | (0.03) | 0.16 | 0.31 | 1.06 |  |  |  |  |  |
| *2 Managerial and technical occupations* |  |  |  |  |  |  |  |  | Ref. |  |  |  |  | 0.07 | -0.15 | 0.15 |  |  |  |  |  |
| *3NM Skilled non-manual occupations* |  |  |  |  |  |  |  |  | 0.02 | (0.11) |  | 0.00 | (0.02) | 0.09 | -0.18 | 0.22 |  |  |  |  |  |
| *3M Skilled manual occupations* |  |  |  |  |  |  |  |  | -0.64 | (0.08) | \*\*\* | -0.09 | (0.01) | 0.05 | -0.75 | -0.53 |  |  |  |  |  |
| *4 Partly-skilled occupations* |  |  |  |  |  |  |  |  | -0.77 | (0.11) | \*\*\* | -0.11 | (0.02) | 0.08 | -0.97 | -0.58 |  |  |  |  |  |
| *5 Unskilled occupations* |  |  |  |  |  |  |  |  | -1.01 | (0.14) |  | -0.14 | (0.02) | 0.12 | -1.29 | -0.73 |  |  |  |  |  |
| CAMSIS (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 0.04 | (0.00) | \*\*\* | 0.00 | (0.00) |
| Intercept | -0.40 | (0.09) | \*\*\* |  |  |  |  |  | -0.63 | (0.08) | \*\*\* |  |  |  |  |  | -2.64 | (0.13) | \*\*\* |  |  |
| Number of observations | 8411 | | | | | | | | 8411 | | | | | | | | 8411 | | | | |
| McFadden’s | 0.38 | | | | | | | | 0.38 | | | | | | | | 0.38 | | | | |
| McFadden’s Adjusted Pseudo | 0.38 | | | | | | | | 0.37 | | | | | | | | 0.38 | | | | |
| Cox-Snell Pseudo | 0.40 | | | | | | | | 0.40 | | | | | | | | 0.40 | | | | |
| Nagelkerke Pseudo | 0.53 | | | | | | | | 0.54 | | | | | | | | 0.54 | | | | |
| Tjur’s | 0.46 | | | | | | | | 0.46 | | | | | | | | 0.46 | | | | |
| AIC | 7009.62 | | | | | | | | 6994.95 | | | | | | | | 6390.51 | | | | |
| BIC | 7087.03 | | | | | | | | 7030.13 | | | | | | | | 6425.69 | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: NCDS [Sweeps 0-4]  Note: Sensitivity Analysis of Social Stratification Measures | | | | | | | | | | | | | | | | | | | | | |

The comparative coefficients for all models are plotted in figure 2.13 to aid in interpretation alongside table 2.29. The log odds and quasi-variance statistics are graphed visually for RGSC and compared to NS-SEC in figure 2.14. From this graph a manual/non-manual divide is evident whereby individuals from non-manual social origins (with exception of RGSC 2) have an increased log odds of continuing schooling over not continuing schooling whereas those individuals from manual social origins have a decreased log odd of continuing schooling. There is a general pattern that emerges from both measures. As the schema decreases from 1.1 to 7 for NS-SEC and 1 to 5 for RGSC, there is a decreased log odds of continuing schooling compared to the reference category. Both log odds and quasi-variance statistics from figure 2.14 illustrate the relative substrative similarities between NS-SEC and RGSC based measures.

A graph with red and blue lines

Description automatically generated

Figure ‎2.13 Coefficient plot comparing all three social stratification models

A diagram of a graph

Description automatically generated with medium confidence

Figure ‎2.14 Comparison of log odds versus quasi-variance statistics of NS-SEC and RGSC measures for NCDS model

Figure 2.15 depicts the predicted probabilities at means of economic activity alongside the average marginal effects of each respective social stratification measure. All graphs are represented using the same common y axis to aid interpretation. Starting with RGSC, with respect to predicted probabilities there is a general monotonic decreased trend for individuals to continue schooling. Moving on to the average marginal effects of RGSC there is a monotonic decrease from RGSC 1 to RGSC 3M, whereby from RGSC 3M to 4 there is a flatlining followed by a small decrease from RGSC 4 to RGSC 5. The largest average increased marginal probability reported in this graph relates to a 9 per cent increased probability of continuing schooling for RGSC 1 compared to RGSC 2. The largest average decreased marginal probability reported in the graph relates to a 11 per cent decreased probability of continuing schooling for RGSC 5 compared to RGSC 2.

Moving on to CAMSIS, figure 2.15 depicts the predicted probabilities at means of continuing schooling alongside the average marginal effects of CAMSIS. The predicted probabilities show a linear monotonic increase for individuals to continue schooling for each point increase in CAMSIS. Average marginal effects flatline are consistently zero.

Figure 2.15 provides a common y axis comparison of each predicted probability and average marginal effects graph for each social stratification measure used within this sensitivity analysis. As interpreted from table 2.25, figure 2.15 demonstrates that NS-SEC and RGSC based models provide substantively identical interpretations, whereas CAMSIS does not.

A group of black and white lines

Description automatically generated

Figure ‎2.15 Comparison of Predictive and AMEs for each social stratification measure for NCDS model

#### Discussion and Conclusion

This investigation of social stratification measures provides two key stories surrounding the use and implementation of various stratification variables within analytical models. The first story relates to the remarkable similarity between NS-SEC and RGSC models. There is across both models substantively identical findings. Both measures find a marked near monotonic decline in continuing schooling for individuals whose social origins are on the lower end of either schema. The RGSC schema, with its fundamental theoretical desire to investigate the manual/non-manual distinction in British society provides a substantive insight that the NS-SEC schema does not. The RGSC model shows a clear divide between the manual and non-manual occupations in British society. NS-SEC’s explicit rejection of the manual/non-manual divide (ISER, 2024) makes it much more difficult to see this sociological trend.

Whilst NS-SEC does not in of itself provide improved statistical associations – or even different statistical associations, it does provide a non a priori definition of class. As well as a way to explore the causal mechanisms more readily than the RGSC measure. The a priori nature of RGSC does provide substantively identical results to the NS-SEC based measure. However, RGSC is not theoretically informed (Rose and Pevalin, 2001). Sociology is not and should not be a data dredging exercise.

There are critiques present of NS-SEC as a measure of social class that also need to be considered. The abandonment of the manual/non-manual divide in British society (Goldthorpe and Hope 1974) was an intentional decision when creating the NS-SEC measure. This was predicated upon a view that the manual/non-manual divided no longer existed to the extent it did in the past (ibid). This is a reasonable argument when discussing contemporary social class structure in Britian. However, when using NS-SEC as a comparative measure between cohorts – especially older, historic datasets, this intentional choice becomes somewhat problematic. British society around the time of the NCDS is characterised by a manual/non-manual division, something that a solely NS-SEC based measure of social stratification finds difficult to capture – at least to the same degree as an RGSC measure. Without using another schema that is sensitive to this divide such as RGSC, an analyst may miss the importance of this divide altogether. Using the NS-SEC schema on its own for historical data prior to its implementation in 2000 may lead to a whitewashing of British socio-historical context. Whilst it would not be appropriate to solely use the RGSC measure on its own considering its lack of theoretical justification, it appears sensible to conduct a sensitivity analysis of both measures when dealing with historical data that predates the NS-SEC schema – and perhaps even after its implementation to see if the manual/non-manual divide truly is dead in Britian.

The remarkable similarity of the RGSC and NS-SEC models, on top of the relative operationalisation similarity between the two social class schemas broaches another important point surrounding the use of social class variables in surveys. There is ultimately an illusion of choice in the datasets used when selecting a social class measure. The lack of inclusion on a variety of measures makes it impossible to construct and compare theoretically distinct social class measures such as Wright’s schema or a manifestation of Bourdieu’s views of social class. Beyond moving towards an inclusion of variables such as wealth, income, education – that are at best epiphenomenal with social class, the options for comparing social class measures is limited. The dominance of one manifestation of social class – in the British case NS-SEC – is damaging to the exploration of social class as a sociological phenomenon in British society and forces social stratification research into a static normal science.

Whilst the RGSC and NS-SEC models are remarkably similar, the CAMSIS model stands alone in comparison. CAMSIS is unlike the other two measures not a measure of social class, it is a measure of social distance. CAMSIS is statistically, and theoretically distinct from social class measures. Whilst CAMSIS and social class measures often have high correlations with one another, a sociological case can be made for the inclusion of both measures within a given analytical model, so long as the theoretical distinction between social distance and social class measures is made. The literature is very careful in distinguishing between social class measures and social scales of distance or prestige, but given their theoretical and substantive differences, they should rightfully be included alongside one another and yet they are not. This lack of inclusion can only be the belief that CAMSIS is capturing the same variance within a model that a social class measure is also capturing. The high correlation between a measure like CAMSIS and a measure like NS-SEC presents issues of collinearity within a model. This is a data driven decision to only include one measure.

This sensitivity analysis was an exploration of different manifestations of social stratification. Each social stratification measure has been statistically and substantively interpreted and critiqued. The result of this section has provided a detailed analysis of different measures of social stratification and their similarities and differences with one another. Overall, this section concludes that when analysing historical datasets like the NCDS employing multiple measures of social class is important to understand the socio-historical context of the time. NS-SEC is a measure that was constructed after the NCDS cohort were born and reflects the reality of its time. RGSC is a measure that was developed on an a priori basis and lacks theoretical rigor. Combined in a sensitivity analysis allows both measures to be studied closely, and any differences to be isolated and investigated. CAMSIS as a measure to be included has to be based upon a theoretical justification that the measure is substantively different from a measure of social class and that the subsequent collinearity can be justified.

### SOC Code Sensitivity analysis using NCDS

Given the sensitivity analysis of social stratification measures, NS-SEC has been selected as the primary model going forward. The sensitivity analysis of social stratification measures provided insight into the construction of different measures and the important of context when dealing with historical datasets like the NCDS. A key component in the construction of all three social stratification measures is the standard occupation classification code or SOC code. This hierarchically organised set of occupations is used when constructing all social stratification variables mentioned hitherto. SOC codes are constructed and organised every 10 years, starting in 1990. Every 10 years the organisation of occupations is reformed through the addition/subtraction of occupations, and through re-organising occupations into different levels. In line with the theme of socio-historical context that has been a constant throughout this thesis, an assessment of SOC codes is a prudent exercise.

Another sensitivity analysis will be conducted comparing the measure of NS-SEC under two different constructions. The first will be NS-SEC constructed using SOC 2000 codes – the base model used previously. The second, will use NS-SEC constructed using SOC 90 codes. These two models will be compared to assess any similarities and differences regarding their substantive effects. Goodness-of-fit statistics will also be assessed to determine the best fit model. A comparison of SOC 2000 and SOC 90 codes for both RGSC and CAMSIS models follows from this initial NS-SEC model to estimate any differences in substantive interpretation if using a different social stratification measure. The purpose of this section is to assess if there is any change in the substantive interpretation of a model if using a different SOC code to construct a given social stratification variable. The SOC codes are meant to resemble the social reality of British occupational structures when they are created. The closer a dataset is to a given SOC code construction, the more likely it is that the dataset reflects the social reality of the occupational structure captured by that given SOC code construction. If a dataset is analysed with the same measure using SOC 90 and SOC 2000 codes and presents different substantive interpretation then there are serious implications for the selection of SOC codes used going forward.

The following tables – 2.26 and 2.27 follow a similar design to those tables produced previously. These tables have been updated with additional information for the SOC 90 measure of NS-SEC as a point of comparison prior to model interpretation. Reflecting first on table 2.26, whilst NS-SEC using the SOC 2000 construction as a difference of 1155.46 deviance from the null, the SOC 90 construction has 847.42 difference in deviance from the null. The statistic is also lower by 0.02 for the SOC 90 construction compared to the SOC 2000 construction. Both AIC and BIC statistics favour the SOC 2000 construction over the SOC 90 construction on its own. Moving on to table 2.27, the full model with SOC 2000 construction of NS-SEC has a 204.94 difference in deviance from the previous model compared to a 146.20 difference in advance for the SOC 90 construction. There is a 0.01 lower statistic for the SOC 90 model over the SOC 2000 model and both AIC and BIC favour the latter over the former. However, the AIC and BIC statistic differences are incredibly small.

Table 2.31 Model building statistics of NS-SEC SOC 90 for NCDS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d. f. (from Null) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 11262.73 | - | - | - | 11264.73 | 11271.77 |
| Null Model + Educational Attainment | 7455.83 | 3,806.90 | 2 | 0.34 | 7459.83 | 7473.91 |
| Null Model + Sex | 11217.31 | 45.42 | 2 | 0.00 | 11221.31 | 11235.38 |
| Null Model + Tenure | 10507.00 | 755.73 | 2 | 0.07 | 10511.00 | 10525.07 |
| Null Model + NS-SEC (SOC 90) | 10415.31 | 847.42 | 8 | 0.08 | 10419.31 | 10433.38 |

Table 2.32 Sequential Model building statistics of NS-SEC SOC 90 for NCDS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 11262.73 | - | - | - | 11264.73 | 11271.77 |
| Null Model + Educational Attainment | 7455.83 | 3,806.90 | 2 | 0.34 | 7459.83 | 7473.91 |
| Null Model + Educational Attainment + Sex | 7400.72 | 55.11 | 2 | 0.34 | 7406.73 | 7427.84 |
| Null Model + Educational Attainment + Sex + Tenure | 7192.56 | 208.16 | 2 | 0.36 | 7200.56 | 7228.71 |
| Null Model + Educational Attainment + Sex + Tenure + NS-SEC (SOC 90) | 7046.36 | 146.20 | 8 | 0.37 | 7068.36 | 7145.77 |

As with the sensitivity analysis that preceded this SOC analysis tables 2.32 and 2.33 provide data on the proportion of total effect size change attributable to rescaling versus mediation of additional variables – comparing SOC constructions of NS-SEC (SOC 2000 in model 1 and SOC 90 in model 2). The variation – if any – provides another useful comparative tool in the difference between SOC constructions of the same social stratification variable.

Both models 2.32 and 2.33 confirm that whilst there are small differences in the total effect size between SOC constructions, the difference in summary statistics is negligible. Both models have very minor rescaling factors and a majority of their total effect size differences originate from mediating effects of additional variables to the model. Somewhat interesting is the slightly diminished confounding ratio and impact that the SOC 90 construction of NS-SEC has upon both educational attainment and to a greater extent, housing tenure. Though this difference is not large enough to warrant further reflection.

Table 2.33 A Comparison of SOC NS-SEC measures using the KHB method

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Model 1 | | Model 2 | |
|  |  | **Log Odds** | **SE** | **Log Odds** | **SE** |
| Educational Attainment | *Reduced* | 3.25\*\*\* | (0.07) | 3.21\*\*\* | (0.07) |
| *Full* | 2.99\*\*\* | (0.06) | 3.02\*\*\* | (0.06) |
| *Difference* | 0.26\*\*\* | (0.03) | 0.20\*\*\* | (0.02) |
| Sex | *Reduced* | -0.50\*\*\* | (0.06) | -0.50\*\*\* | (0.06) |
| *Full* | -0.50\*\*\* | (0.06) | -0.50\*\*\* | (0.06) |
| *Difference* | 0.00 | (0.02) | 0.00 | (0.01) |
| Housing Tenure | *Reduced* | -0.88\*\*\* | (0.06) | -0.88\*\*\* | (0.06) |
| *Full* | -0.63\*\*\* | (0.06) | -0.68\*\*\* | (0.06) |
| *Difference* | -0.25\*\*\* | (0.03) | -0.20\*\*\* | (0.02) |

Table 2.34 KHB Summary statistics Comparing SOC NS-SEC models

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Model 1 | | | Model 2 | | |
|  | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** |
| Educational Attainment | 1.09 | 8.03 | 1.04 | 1.07 | 6.07 | 1.03 |
| Sex | 0.99 | -0.78 | 1.04 | 1.00 | -0.00 | 1.03 |
| Housing Tenure | 1.40 | 28.38 | 1.01 | 1.30 | 22.95 | 1.01 |

Table 2.34 displays a comparison of the full proposed model using the SOC 2000 construction of NS-SEC in one model and the SOC 90 construction of NS-SEC in the second model. Log odds coefficients, average marginal effects, and quasi-variance statistics are provided for ease of interpretation. Unsurprisingly educational attainment, sex, and housing tenure do not deviate substantively. Whilst the SOC 2000 construction finds NS-SEC 1.2, and 3-7 statistically significant, the SOC 90 construction only finds NS-SEC 4-7 statistically significant. This may be due to how occupations in SOC 90 were re-adjusted in SOC 2000. Across NS-SEC 4-7 the substantive significance is identical across all measures though SOC 90 has a small reduction (around 1 per cent) in the average marginal effects.

Table 2.35 Comparison of SOC measures for NS-SEC for NCDS model

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | NS-SEC (SOC 2000) | | | Average Marginal Effects | | Quasi-variance | | | NS-SEC (SOC 90) | | | Average Marginal Effects | | Quasi-variance | | |
| Economic Activity: ‘Don’t Continue Schooling’ Reference Category | Coef. | S.E. | Sig. | Prob. | S.E. | S.E. | LB | UB | Coef. | S.E. | Sig. | Prob. | S.E. | S.E. | LB | UB |
| Continue Schooling |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Less than five O’levels | Ref. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Five or More O’levels | 2.98 | (0.07) | \*\*\* | 0.59 | (0.01) |  |  |  | 3.02 | (0.06) | \*\*\* | 0.60 | (0.01) |  |  |  |
| Sex |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Female | Ref. | (.) |  | (.) | (.) |  |  |  | (.) | (.) |  | (.) | (.) |  |  |  |
| Male | -0.50 | (0.06) | \*\*\* | -0.06 | (0.01) |  |  |  | -0.49 | (0.06) | \*\*\* | -0.06 | (0.01) |  |  |  |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Own Home | Ref. | (.) |  | (.) | (.) |  |  |  | (.) | (.) |  | (.) | (.) |  |  |  |
| Do not Own Home | -0.63 | (0.06) | \*\*\* | -0.08 | (0.01) |  |  |  | -0.68 | (0.06) | \*\*\* | -0.09 | (0.01) |  |  |  |
| NS-SEC |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1.1 | 0.02 | (0.19) |  | 0.00 | (0.03) | 0.17 | -0.36 | 0.40 | -0.17 | (0.91) |  | -0.03 | (0.14) | 0.91 | -2.19 | 1.84 |
| 1.2 | 0.48 | (0.17) | \*\*\* | 0.08 | (0.03) | 0.15 | 0.14 | 0.82 | 0.25 | (0.19) |  | 0.04 | (0.03) | 0.16 | -0.11 | 0.61 |
| 2 | Ref. |  |  |  |  | 0.09 | -0.19 | 0.19 |  |  |  |  |  | 0.11 | -0.24 | 0.24 |
| 3 | -0.25 | (0.13) | \* | -0.04 | (0.02) | 0.09 | -0.46 | -0.04 | -0.19 | (0.14) |  | -0.03 | (0.02) | 0.09 | -0.39 | 0.02 |
| 4 | -0.89 | (0.12) | \*\*\* | -0.13 | (0.02) | 0.09 | -1.08 | -0.70 | -0.85 | (0.15) | \*\*\* | -0.12 | (0.02) | 0.11 | -1.09 | -0.62 |
| 5 | -0.76 | (0.11) | \*\*\* | -0.11 | (0.02) | 0.07 | -0.93 | -0.60 | -0.75 | (0.13) | \*\*\* | -0.11 | (0.02) | 0.07 | -0.91 | -0.60 |
| 6 | -0.89 | (0.11) | \*\*\* | -0.13 | (0.02) | 0.07 | -1.05 | -0.72 | -0.80 | (0.12) | \*\*\* | -0.12 | (0.02) | 0.06 | -0.93 | -0.66 |
| 7 | -1.11 | (0.11) | \*\*\* | -0.16 | (0.02) | 0.07 | -1.26 | -0.95 | -1.08 | (0.12) | \*\*\* | -0.15 | (0.02) | 0.06 | -1.22 | -0.94 |
| Intercept | -0.40 | (0.09) | \*\*\* |  |  |  |  |  | -0.35 | (0.11) | \*\*\* |  |  |  |  |  |
| Number of observations | 8411 | | | | | | | | 8411 | | | | | | | |
| McFadden’s | 0.38 | | | | | | | | 0.37 | | | | | | | |
| McFadden’s Adjusted Pseudo | 0.37 | | | | | | | | 0.37 | | | | | | | |
| Cox-Snell Pseudo | 0.40 | | | | | | | | 0.40 | | | | | | | |
| Nagelkerke Pseudo | 0.53 | | | | | | | | 0.53 | | | | | | | |
| Tjur’s | 0.46 | | | | | | | | 0.46 | | | | | | | |
| AIC | 7009.62 | | | | | | | | 7068.36 | | | | | | | |
| BIC | 7087.03 | | | | | | | | 7145.77 | | | | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: NCDS [Sweeps 0-4]  Note: Sensitivity Analysis of Social Stratification Measures | | | | | | | | | | | | | | | | |

Both the SOC 90 and SOC 2000 model share near identical measures. Each models log odds and quasi-variance statistics are graphed in figures 2.17. From figure 2.17 the standard errors from NS-SEC 1.1 are inflated compared to other NS-SEC categories. This is primarily driven through the small n of individuals within NS-SEC 1.1 under a SOC 90 construction. Using figure 2.17 a direct graphical comparison is made between SOC constructions. Whilst the same overall substantive trend can be made using both constructions, the inflated standard errors in NS-SEC 1.1 for SOC 90 constructions is the most obvious deviation between the two models.

Inflated standard errors once again can be seen in figure 2.18 with a direct comparison of predicted probabilities and average marginal effects for SOC 90 and SOC 2000 constructions of NS-SEC. Both predictive margins and average marginal effects are identical, though the confidence intervals for the SOC 90 model are substantially larger than the SOC 2000 model.

A screenshot of a graph

Description automatically generated

Figure ‎2.17 Comparison of log odds and quasi-variance statistics for NS-SEC SOC codes for NCDS model

A screenshot of a computer screen

Description automatically generated

Figure ‎2.18 Comparison of Predictive and AMEs of NS-SEC SOC Codes for NCDS model

Moving on to the comparison of other social stratification measures, starting with RGSC. Table 2.35 details the model statistics for RGSC SOC 90. The SOC 90 construction of RGSC has a reduction of 871.21 deviance from the null model. The SOC 2000 construction of RGSC detailed in table 2.30 shows RGSC has as reduction of 1026.66 deviance from the null model. Table 2.36 details that the SOC 90 construction of RGSC provides a reduction of 150.75 from the previous model whereas the SOC 2000 construction of RGSC detailed in table 2.24 has a 167.68 reduction in deviance from the prior model. The statistics for both the SOC 90 and SOC 2000 constrtuctions of RGSC models are virtually identical.

Table 2.36 Model Statistics of RGSC SOC 90 for NCDS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d. f. (from Null) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 11262.73 | - | - | - | 11264.73 | 11271.77 |
| Null Model + Educational Attainment | 7455.83 | 3,806.90 | 2 | 0.34 | 7459.83 | 7473.91 |
| Null Model + Sex | 11217.31 | 45.42 | 2 | 0.00 | 11221.31 | 11235.38 |
| Null Model + Tenure | 10507.00 | 755.73 | 2 | 0.07 | 10511.00 | 10525.07 |
| Null Model + RGSC (SOC 90) | 10391.52 | 871.21 | 6 | 0.07 | 10403.52 | 10445.75 |

Table 2.37 Sequential Model Statistics of RGSC SOC 90 for NCDS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 11262.73 | - | - | - | 11264.73 | 11271.77 |
| Null Model + Educational Attainment | 7455.83 | 3,806.90 | 2 | 0.34 | 7459.83 | 7473.91 |
| Null Model + Educational Attainment + Sex | 7400.72 | 55.11 | 2 | 0.34 | 7406.73 | 7427.84 |
| Null Model + Educational Attainment + Sex + Tenure | 7192.56 | 208.16 | 2 | 0.36 | 7200.56 | 7228.71 |
| Null Model + Educational Attainment + Sex + Tenure + RGSC (SOC 90) | 7041.81 | 150.75 | 6 | 0.38 | 7059.81 | 7123.14 |

As produced earlier for SOC constructions of NS-SEC, the KHB method is employed to compare SOC constructions of RGSC and the variation in total effects measured against rescaling versus mediating effects of additional variables added. Model 1 represents RGSC SOC 2000 and model 2 represents RGSC SOC 90.

Similar to the findings of the comparison of SOC constructions for NS-SEC, the RGSC models report similar rescaling and confounding statistics. The change in total effects is very small. Whilst similar to the NS-SEC models, the RGSC SOC constructions also see a slight overall reduction in the confounding percentage and ratio of the SOC 90 model – the substantive nature of this is minor, even less than the NS-SEC phenomena.

Table 2.38 A Comparison of SOC RGSC measures using the KHB method

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Model 1 | | Model 2 | |
|  |  | **Log Odds** | **SE** | **Log Odds** | **SE** |
| Educational Attainment | *Reduced* | 3.22\*\*\* | (0.07) | 3.22\*\*\* | (0.07) |
| *Full* | 3.00\*\*\* | (0.06) | 3.02\*\*\* | (0.06) |
| *Difference* | 0.22\*\*\* | (0.02) | 0.19\*\*\* | (0.02) |
| Sex | *Reduced* | -0.50\*\*\* | (0.06) | -0.50\*\*\* | (0.06) |
| *Full* | -0.50\*\*\* | (0.06) | -0.50\*\*\* | (0.06) |
| *Difference* | 0.01 | (0.02) | -0.00 | (0.01) |
| Housing Tenure | *Reduced* | -0.88\*\*\* | (0.06) | -0.88\*\*\* | (0.06) |
| *Full* | -0.64\*\*\* | (0.06) | -0.70\*\*\* | (0.06) |
| *Difference* | -0.24\*\*\* | (0.03) | -0.18\*\*\* | (0.02) |

Table 2.39 KHB Summary statistics Comparing SOC RGSC models

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Model 1 | | | Model 2 | | |
|  | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** |
| Educational Attainment | 1.07 | 6.78 | 1.03 | 1.06 | 6.01 | 1.03 |
| Sex | 0.99 | -1.35 | 1.03 | 1.01 | 0.55 | 1.03 |
| Housing Tenure | 1.37 | 27.06 | 1.01 | 1.26 | 20.67 | 1.01 |

The comparison of SOC 2000 versus SOC 90 constructions of RGSC models are provided in table 2.39. Log odds, average marginal effects, and quasi-variance statistics are provided. Like the NS-SEC comparisons, the RGSC models provide near identical substantive interpretations for all other analytical variables. Whilst there is some general agreement across each RGSC model where both share overlapping statistical significance – for example, RGSC 4 for both models provides a near identical substantive interpretation. There are however some discrepancies – more prominent in RGSC 3M, whilst both models report a decrease in percentage points continuing schooling for members of RGSC 3M compared to RGSC 2, the size of this effect is substantially different across models. For the SOC 2000 model, the average marginal effect is -9 percentage points, for the SOC 90 model, it is -16 percentage points. On top of this, whilst the SOC 2000 model finds RGSC 1 to be statistically significant, the SOC 90 model does not. The RGSC model also finds RGSC 3NM and 5 to be statistically significant whilst SOC 2000 does not. Overall, whilst there are some general similarities between the two models, the overarching substantive pattern is different dependent on the SOC construction used for analysis. RGSC appears to be more sensitivity to SOC constructions under this model than other social stratification measures such as NS-SEC.

Table 2.40 Comparison of RGSC SOC for NCDS Model

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | RGSC (SOC 2000) | | | Average Marginal Effects | | Quasi-variance | | | RGSC (SOC 90) | | | Average Marginal Effects | | Quasi-variance | | |
| Economic Activity: ‘Don’t Continue Schooling’ Reference Category | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **S.E.** | **LB** | **UB** | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **S.E.** | **LB** | **UB** |
| Continue Schooling |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Five or More O’levels* | 3.00 | (0.06) | \*\*\* | 0.60 | (0.01) |  |  |  | 3.02 | (0.06) | \*\*\* | 0.60 | (0.01) |  |  |  |
| Sex |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. |  |  |  |  |  |  |  | (.) | (.) |  | (.) | (.) |  |  |  |
| *Male* | -0.50 | (0.06) | \*\*\* | -0.06 | (0.01) |  |  |  | -0.49 | (0.06) | \*\*\* | -0.06 | (0.01) |  |  |  |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. |  |  |  |  |  |  |  | (.) | (.) |  | (.) | (.) |  |  |  |
| *Do not Own Home* | -0.64 | (0.06) | \*\*\* | -0.09 | (0.01) |  |  |  | -0.70 | (0.06) | \*\*\* | -0.10 | (0.01) |  |  |  |
| RGSC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1* | 0.68 | (0.18) | \*\*\* | 0.11 | (0.03) | 0.16 | 0.31 | 1.06 | 0.26 | (0.21) |  | 0.04 | (0.03) | 0.17 | -0.14 | 0.66 |
| *2* | Ref. |  |  |  |  | 0.07 | -0.15 | 0.15 | Ref. |  |  |  |  | 0.11 | -0.26 | 0.26 |
| *3NM* | 0.02 | (0.11) |  | 0.00 | (0.02) | 0.09 | -0.18 | 0.22 | -0.38 | (0.14) | \*\* | -0.06 | (0.02) | 0.08 | -0.56 | -0.19 |
| *3M* | -0.64 | (0.08) | \*\*\* | -0.09 | (0.01) | 0.05 | -0.75 | -0.53 | -1.12 | (0.13) | \*\*\* | -0.16 | (0.02) | 0.06 | -1.25 | -0.99 |
| *4* | -0.77 | (0.11) | \*\*\* | -0.11 | (0.02) | 0.08 | -0.97 | -0.58 | -0.86 | (0.12) | \*\*\* | -0.13 | (0.02) | 0.05 | -0.97 | -0.74 |
| *5* | -1.01 | (0.14) |  | -0.14 | (0.02) | 0.12 | -1.29 | -0.73 | -1.02 | (0.15) | \*\*\* | -0.15 | (0.02) | 0.10 | -1.26 | -0.78 |
| Intercept | -0.63 | (0.08) | \*\*\* |  |  |  |  |  | -0.26 | (0.12) | \* |  |  |  |  |  |
| Number of observations | 8411 | | | | | | | | 8411 | | | | | | | |
| McFadden’s | 0.38 | | | | | | | | 0.38 | | | | | | | |
| McFadden’s Adjusted Pseudo | 0.37 | | | | | | | | 0.37 | | | | | | | |
| Cox-Snell Pseudo | 0.40 | | | | | | | | 0.37 | | | | | | | |
| Nagelkerke Pseudo | 0.54 | | | | | | | | 0.52 | | | | | | | |
| Tjur’s | 0.46 | | | | | | | | 0.46 | | | | | | | |
| AIC | 6994.95 | | | | | | | | 7059.81 | | | | | | | |
| BIC | 7030.13 | | | | | | | | 7123.14 | | | | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: NCDS [Sweeps 0-4]  Note: Sensitivity Analysis of Social Stratification Measures | | | | | | | | | | | | | | | | |

Log odds and quasi-variance statistics are detailed graphically for NS-SEC in figure 2.20. The SOC 90 construction of RGSC demonstrates a clear manual divide amongst members of RGSC 3M-5 in comparison to the reference category of RGSC 2. Unlike the SOC 2000 construction of RGSC detailed in figure 2.20, there are no statistically significant differences between RGSC 2 and other members of non-manual RGSC categories. Both RGSC measures make it clear that in comparison to RGSC 2, manual occupations are statistically and substantively different in continuing schooling. This presents a clear manual/non-manual divide. Figure 2.21 details the predictive probabilities and average marginal effects of both RGSC measures. Like the story told previously, the overlapping statistical significance of RGSC occupations tells a similar substantive story.

A screenshot of a computer screen

Description automatically generated

Figure ‎2.20 Comparison of log odds versus quasi-variance statistics for RGSC SOC Codes

A screenshot of a computer screen

Description automatically generated

Figure ‎2.21 Comparison of Predictive and AMEs for RGSC SOC Codes for NCDS Model

Finally, now moving on to the CAMSIS SOC 90 measure of social stratification. Table 2.40 details the model statistics for CAMSIS SOC 90. The SOC 90 construction of CAMSIS has a reduction of 780.65 deviance from the null model. Table 2.41 details that the SOC 90 construction of CAMSIS provides a reduction of 113.09 from the previous model. The statistics for both the SOC 90 and SOC 2000 constrtuctions of CAMSIS models are virtually identical.

Table 2.41 Model Statistics of CAMSIS SOC 90 for NCDS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d. f. (from Null) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 11262.73 | - | - | - | 11264.73 | 11271.77 |
| Null Model + Educational Attainment | 7455.83 | 3,806.90 | 2 | 0.34 | 7459.83 | 7473.91 |
| Null Model + Sex | 11217.31 | 45.42 | 2 | 0.00 | 11221.31 | 11235.38 |
| Null Model + Tenure | 10507.00 | 755.73 | 2 | 0.07 | 10511.00 | 10525.07 |
| Null Model + CAMSIS (SOC 90) | 10482.08 | 780.65 | 2 | 0.07 | 10486.08 | 10500.16 |

Table 2.42 Sequential Model Statistics of CAMSIS SOC 90 for NCDS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 11262.73 | - | - | - | 11264.73 | 11271.77 |
| Null Model + Educational Attainment | 7455.83 | 3,806.90 | 2 | 0.34 | 7459.83 | 7473.91 |
| Null Model + Educational Attainment + Sex | 7400.72 | 55.11 | 2 | 0.34 | 7406.73 | 7427.84 |
| Null Model + Educational Attainment + Sex + Tenure | 7192.56 | 208.16 | 2 | 0.36 | 7200.56 | 7228.71 |
| Null Model + Educational Attainment + Sex + Tenure + CAMSIS (SOC 90) | 7079.47 | 113.09 | 2 | 0.37 | 7089.47 | 7124.66 |

Finally, as with the other two social stratification measures the KHB method is employed to compare the total effect sizes compared with rescaling versus mediating effects of additional variables. Model 1 represents CAMSIS SOC 2000 and model 2 represents CAMSIS SOC 90. The total effect size differences reported in table 2.42 are like prior SOC model comparisons small. Table 2.43 does detail the recurrent phenomena that the SOC 90 construction of the social stratification variable produces a recued confounding percentage and ratio compared to its SOC 2000 counterpart. However, both SOC constructions report near identical rescaling factors. All social stratification measures regardless of their SOC construction report that rescaling has a small impact of the change of total effect size in analytical models reported. The comparisons between models using the same sample can be made confidently knowing that the rescaling factor is negligible.

Table 2.43 A Comparison of SOC CAMSIS measures using the KHB method

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Model 1 | | Model 2 | |
|  |  | **Log Odds** | **SE** | **Log Odds** | **SE** |
| Educational Attainment | *Reduced* | 3.25\*\*\* | (0.07) | 3.19\*\*\* | (0.06) |
| *Full* | 2.97\*\*\* | (0.06) | 3.03\*\*\* | (0.06) |
| *Difference* | 0.27\*\*\* | (0.03) | 0.16\*\*\* | (0.02) |
| Sex | *Reduced* | -0.50\*\*\* | (0.06) | -0.49\*\*\* | (0.06) |
| *Full* | -0.51\*\*\* | (0.06) | -0.49\*\*\* | (0.06) |
| *Difference* | 0.01 | (0.02) | 0.00 | (0.01) |
| Housing Tenure | *Reduced* | -0.88\*\*\* | (0.06) | -0.88\*\*\* | (0.06) |
| *Full* | -0.60\*\*\* | (0.06) | -0.72\*\*\* | (0.06) |
| *Difference* | -0.29\*\*\* | (0.03) | -0.16\*\*\* | (0.02) |

Table 2.44 KHB Summary statistics Comparing SOC CAMSIS models

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Model 1 | | | Model 2 | | |
|  | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** |
| Educational Attainment | 1.09 | 8.45 | 1.04 | 1.05 | 5.15 | 1.02 |
| Sex | 0.97 | -2.75 | 1.03 | 1.00 | -0.27 | 1.02 |
| Housing Tenure | 1.48 | 32.45 | 1.01 | 1.23 | 18.37 | 1.01 |

The comparison of SOC 2000 versus SOC 90 constructions of CAMSIS models are provided in table 2.44. Log odds and average marginal effects statistics are provided. Like both NS-SEC and RGSC models other analytical variables included in the model are substantively identical between SOC measures. Both SOC 2000 and SOC 90 constructions of CAMSIS provide near identical log odds (a 0.01 difference between the measures), identical statistical and substantive significance. There appears to be no difference between using the SOC 2000 or SOC 90 measure of CAMSIS. Compared to other social stratification measures like RGSC, where there are some slight deviations between SOC constructions this appears to suggest that measures of social class such as NS-SEC and especially RGSC are more sensitive to changing SOC classifications compared to a social distance scale such as CAMSIS.

Table 2.45 Comparison of CAMSIS SOC Codes for NCDS Model

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | CAMSIS (SOC 2000) | | | Average Marginal Effects | | CAMSIS (SOC 90) | | | Average Marginal Effects | |
| Economic Activity: ‘Don’t Continue Schooling’ Reference Category | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** |
| Continue Schooling |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. |  |  |  |  |  |  |  |  |  |
| *Five or More O’levels* | 2.97 | (0.06) | \*\*\* | 0.59 | (0.01) | 3.03 | (0.06) | \*\*\* | 0.60 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. |  |  |  |  |  |  |  |  |  |
| *Male* | -0.51 | (0.06) | \*\*\* | -0.07 | (0.01) | -0.49 | (0.06) | \*\*\* | -0.06 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. |  |  |  |  |  |  |  |  |  |
| *Do not Own Home* | -0.59 | (0.06) | \*\*\* | -0.08 | (0.01) | -0.72 | (0.06) | \*\*\* | -0.10 | (0.01) |
| CAMSIS (SOC 2000) | 0.04 | (0.00) | \*\*\* | 0.00 | (0.00) | 0.03 | (0.00) | \*\*\* | 0.00 | (0.00) |
| Intercept | -2.64 | (0.13) | \*\*\* |  |  | -2.15 | (0.13) | \*\*\* |  |  |
| Number of observations | 8411 | | | | | 8411 | | | | |
| McFadden’s | 0.38 | | | | | 0.38 | | | | |
| McFadden’s Adjusted Pseudo | 0.38 | | | | | 0.37 | | | | |
| Cox-Snell Pseudo | 0.40 | | | | | 0.37 | | | | |
| Nagelkerke Pseudo | 0.54 | | | | | 0.52 | | | | |
| Tjur’s | 0.46 | | | | | 0.46 | | | | |
| AIC | 6390.51 | | | | | 7089.47 | | | | |
| BIC | 6425.69 | | | | | 7124.66 | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: NCDS [Sweeps 0-4]  Note: Sensitivity Analysis of Social Stratification Measures | | | | | | | | | | |

Figure 2.23 details the predictive probabilities and average marginal effects of both CAMSIS measures – there is an identical substantive pattern that emerges. Though the SOC 2000 construction of CAMSIS appears to have a slightly steeper predictive margins gradient compared to the SOC 90 construction of CAMSIS.

A graph of a function

Description automatically generated with medium confidence

Figure ‎2.23 Comparison of Predictive and AMEs for CAMSIS SOC Codes

Figure 2.24 provides a direct comparison of all model’s coefficients. This is directly compared in figure 2.25 to the SOC 2000 constructions of the same measures in the models to provide comparison for all six models directly.

Figure 2.26 details all three social stratification measures by SOC 90 construction. Except for inflated CIs across each measure, the same substantive story corroborates that of the previous chapter. NS-SEC measures are remarkably similar in their substantive interpretation, RGSC has some slight deviations but provides generally similar conclusions and CAMSIS demonstrates a lack of substantive significance. Figure 2.27 attempts to combine all sections of the NCDS analysis so far by plotting predictive probabilities and average marginal effects for each social stratification measure and each SOC construction.

A diagram of a graph

Description automatically generated with medium confidence

Figure ‎2.24 Coefficient plot comparing all SOC 90 models

A screenshot of a computer screen

Description automatically generated

Figure ‎2.25 Comparative Coefficient plots by SOC constructions of social stratification measures

A group of rectangular shapes

Description automatically generated with medium confidence

Figure ‎2.26 Comparison of Predictive and AMEs for all social stratification measures for NCDS model for SOC 90 codes

A group of graphs on a white background

Description automatically generated

Figure ‎2.27 Comparison of Predictive and AMEs of all SOC codes for NCDS Mode

#### Discussion and Conclusions

There are two key points to be made from this section. The first is that all SOC 90 constructions have inflated CIs compared to their SOC 2000 counterparts when analyzing NCDS data. The second is that whilst there is a general pattern of substantive significance that echoes across SOC measures, there are some marginal differences between the SOC 2000 constructions of social class measures (NS-SEC and RGSC) compared to their SOC 90 counterparts. This differences are much more pronounced within RGSC constructions compared to NS-SEC constructions. This is not true for CAMSIS. This appears to suggest that social class measures of social stratification are more sensitive to the changes made within SOC classifications compared to other measures of social stratification, in this instance a measure of social distance that also uses SOC classifications. CAMSIS could then, be considered a robust measure to implement when studying large timeframes with altering occupational patterns within British society. Given its different substantive interpretation to the two social class measures, however, this recommendation is caveated with the belief that social class measures should be used alongside any introduction of CAMSIS.

### Handling Missing Data

#### Missing Data

Missing data is an essential component of any longitudinal data analysis – the primary concern being that missing data and non-response are bound to affect the inferences made by the analysis of longitudinal studies (Hawkes and Plewis, 2006: 479; Silverwood *et al.*, 2021). The various factors that account for sample attrition in the datasets outlined in chapter 2.3 have the potential to present real issues as they relate to comprehensive data analysis. For analysis, those who exit the sample due to death or emigration are considered ‘natural’ from the original sample. Those, however, that either cannot be found, reject continued participation, etc., are individuals that we hold partial data on – using this partial data within an analysis could be beneficial.

There are three primary types of classification when dealing with missing data. For ease of interpretation, suppose that only one variable Y has missing data, and that another set of variables represented by the vector X, is always observed (Marsden and Wright, 2010). The data is MCAR if the probability that Y is missing does not depend on either X or Y itself. Evaluating the assumption that missingness on Y depends on some observed variable in X is straightforward. Allison uses the example of income depending on gender by testing whether the proportions of men and women who report their income differ – a logistic regression in which the dependent variable is the response indicator could be estimated and significant coefficients would suggest a violation of the MCAR mechanism (ibid). Testing whether missingness on Y does not depend on Y itself is much more complicated. Unless we have existing linked data such as tax records in the income example, it is almost impossible to evaluate this assumption.

The second missingness mechanism is missing at random (MAR). Data on Y is considered MAR if the probability that Y is missing does not depend on Y, once we control for X. MAR allows for missingness on Y to depend on other variables so long as it does not depend on Y itself.

Finally, missing not at random (MNAR) means missingness depends on unobserved values (Silverwood et al. 2021), and that the probability that Y is missing depends on Y itself, after adjusting for X (Marsden and Wright, 2010). For example, people who have been arrested may be less likely to report their arrest status.

If data is found to be MAR or MCAR, then approaches like multiple imputation (MI), Full Information Maximum likelihood (FIML), and inverse probability weighting (IPW) are made available – the former being extensively documented with the NCDS (Hawkes and Plewis 2006). These ‘gold standard’ approaches to handling missing data have also been found to produce optimal estimates in the MNAR case but it is difficult to have confidence that any given MNAR model is correct (Marsden and Wright, 2010).

When dealing with missing data, there are multiple methods to tackle the problem ranging from an ‘inadequate’ to ‘gold’ standard. The first is listwise deletion. Listwise deletion removes all observations from the data with a missing value in one or more of the variables included in the analysis. This is also known as Complete Records Analysis (CRA). The CRA approach is unpredictable; there is no way to know the consequences of this loss of information if data is found to be MAR (Carpenter and Kenward, 2012). When data is found to be MAR, a CRA approach is inadequate at handling missing data.

Depending on the variable (either metric or categorical) a simple approach to handling missing data would be to use a single mean or single modal imputation. This in the example of a categorical variable takes the mode of the value in said variable and imputes that modal value across all missing values in the data. Single imputation ignores all uncertainty and always underestimates the variance in each model. Advocates of this approach argue that whilst not perfect this approach doesn’t delete a single case and incorporates all available information into a given model. However, this method does not have any confidence in its results. There is a possibility that the estimates from this method may fall close to the true range, of course the exact opposite is equally likely. The use of single use imputation has been consistently and conclusively shown to perform poorly except under exceptionally special conditions (Collins, Schafer and Kam, 2001; Little and Rubin, 2019). For these reasons, single use imputation is an inadequate method to handle missing data.

Dummy variable adjustment is another method of handling missing data. Dummy variable adjustment may appear to be in the same category of handling missing data methods as single use imputation. This is, however, not the case. Dummy variable adjustment is where all missingness at the given variable is coded to a value within the model. In the example of a binary dummy variable, all missingness is coded to either equal zero or equal one. This does have the identical appeal to single use imputation of deleting no cases and incorporates all information into the regression model. However, there is a substantive difference between the two techniques. For the simple model of data missing at Y variable, a dummy variable adjustment will not provide the ‘true’ estimates but if the complete records analysis is compared to a model where all missingness equals zero and another model where all missingness equals one, then the range of the estimates can be located. Whilst Jones (1996) demonstrated that dummy variable adjustment yields biased parameter estimates even when the data is MCAR, the ability to provide a range of the estimates does provide some utility to this technique. Given a MAR example where the reported estimates are a reduced form from their ‘true’ values, iff the complete case analysis and both dummy variable adjustment models present a beta coefficient that is throughout all models positive, one can present those results like how we ought to interpret log odds. The results would present evidence for a positive coefficient – though the exact size is unknown, some information can be gathered. For this reason, dummy variable adjustment provides some utility in certain missing data scenarios. This technique has most utility in scenarios where missingness is so great that it begins to stretch the abilities of even gold standard techniques. This method for handling missing data is not perfect, but it does provide utility and allows the use of data that has large amounts of missingness.

Another method that deals with missing data is the use of survey weights. Survey weights consider missingness. Inverse Probability Weighting (IPW) creates weighted copies of complete records to remove selection bias introduced by missing data. Whilst IPW is a method of dealing with missing data, alternatives such as multiple imputation are regarded as much more efficient as IPW only determine weights from incomplete cases and partially observed cases are discarded int eh weighted analysis. Due to this, weighted estimates can have unacceptably high variance (Seaman *et al.*, 2012; Seaman and White, 2013; Little, Carpenter and Lee, 2022).

There are two ‘gold standard’ approaches to handling missing data, Multiple Imputation (MI) and Maximum Likelihood (ML). Referring to the latter method first, there are currently three ML estimation algorithms for use when missing data is present with either an MCAR or MAR mechanism. The first is the multiple-group method, whereby a sample is divided into subgroups which each share the same pattern of missing data. A likelihood function is computed for each of the subgroups and the groupwise likelihood functions are accumulated across the entire sample and maximised. There are some practical issues of implementing this multiple-group based ML approach (Enders, 2001). The major drawback of this approach however is that it is a group level, rather than individual level ML estimation. Another ML estimation is the expectation-maximisation (EM). This estimation uses a two-step iterative procedure where missing observations are filled in or imputed and the unknown parameters are estimated using maximum likelihood missing data algorithms. The EM approach can only be used to obtain ML estimates of a mean vector and covariance matrix and as a result standard errors will be negatively biased and bootstrapping is recommended (Enders, 2001). The final ML approach discussed here is the Full Information Maximum Likelihood (FIML) estimation. It has also been called the raw maximum likelihood estimation for its likelihood function being calculated at the individual. It is also exceptionally easy to implement compared to the other estimation procedures discussed (Enders, 2001). For these reasons, going forward ML discussions of handling missing data specifically refer to the FIML approach rather than the multiple-group or EM approach.

Multiple Imputation is the second of the ‘Gold standard’ handling missing data methods. Multiple imputation generates replacement values or imputations for the missing data values and repeats this procedure over many iterations to produce a ‘semi-Bayesian’ framework for the most appropriate fit of estimates. For multiple imputation models to be compared to a complete records analysis, the former needs to be ‘‘congenial’’ (White, Royston and Wood, 2011) with the latter. Congeniality or consistency in this respect means that the same variables in the complete record analysis are identical to those included in multiple imputation. Suppose the variables between complete records analysis and multiple imputation models differ. In that case, the correct variance/covariance matrix will not be estimated, and a substantive comparison between the two will become impossible and impracticable due to a loss of statistical power (Von Hippel, 2009; Lynch and Von Hippel, 2013).

Multivariate imputation by chained equations (MICE) is a form of multiple imputation that fills in or imputes missing data within a given dataset through iterative predictive models or *k* imputations. This specification is required when imputing a variable that must only take on specific values, such as the categorical nature of the economic activity response variable within the current analytical model. Using MICE, each imputation *k* is drawn from the posterior distribution of the parameters in the given imputation model, and then the model itself is imputed (Carpenter and Kenward, 2012). To create the *k*th imputation, new parameters are drawn from the posterior distribution. Multiple Imputation following MICE draws from Bayesian influences on the distribution of missing data upon observed data. An essential advantage of Multiple Imputation is that it can be applied for data missing at the response variable or its covariates (Carpenter and Kenward, 2012).

Multiple imputation uses auxiliary variables – variables not included in the main model but are used when setting the data to be imputed. The auxiliary variables main function is to improve the predictive ability of the imputation model over and above the information recovered from just using information provided by the analytical variables in the model (Collins, Schafer and Kam, 2001). Auxiliary variables are essential when there are high levels of missingness upon a given variable (Johnson and Young, 2011; Young and Johnson, 2011). There is no strict threshold for what an auxiliary variable needs to be included within the imputation; however, some have recommended an r > 0.4 on at least one of the analytical variables within the model (Allison, 2012a). However, this is disputed (Enders, 2010). Others, such as Silverwood et al. (2021), argue that if an auxiliary variable is predictive of the outcome variable, it makes them suitable for inclusion within the imputation model. An auxiliary variable does not have the requirement that the given variable has to have complete information to be valuable – auxiliary variables can still be influential when they have missingness (Enders, 2010).

Multiple Imputation can be implemented easily and readily across software platforms unlike FIML. Multiple imputation does however have some drawbacks. It can be a lengthy procedure that has the potential to induce human error due to the need to select auxiliary variables, set the correct data for imputation, and set the correct seed etc. There is also a time efficiency argument, whereby for multiple imputation, if the dataset is large, or there are large amounts of missingness, then the time to impute the model of interest can take a large amount of time. MI is an attractive method because it is practical and widely applicable (Carpenter and Kenward, 2012).

Whilst original literature on missing data and MI typically referred to large datasets with marginal levels of missingness, contemporary studies and simulations have increasingly stretched and stress tested the limits of MI (Hardt *et al.*, 2013). A simulation by Hardt et al (2013) demonstrated that large amounts of missingness can be present within a model without breaking down MI or FIML mechanisms (ibid). Whilst their simulation stops at n=200 where 40 per cent missingness is acceptable, there general argument is that the greater the n the larger the missingness can be within a model without breaking MI or FIML so long as the models themselves are appropriately specified. Imputation based models are consistently found to outperform a CRA in both absolute bias and Root Mean Squared Error (RMSE) with increasing levels of missingness (Hyuk Lee and Huber Jr., 2021). The most extreme case from Madely-Down et al (2019) demonstrates that so long as the imputation model is properly specified and data are MAR then unbiased results can be obtained even with up to 90 per cent missingness. An imputation model compared to a CRA can achieve a reduction in 99.97 per cent bias when missingness is at 90 per cent (ibid).

When dealing with MI, the subsequent question that naturally follows is how many imputations are sufficient? Silverwood et al. (2021) suggest that anything around 50 imputations would be sufficient for reliable estimation of the point estimate and estimating p-values with little error. Sometimes, with large samples with sizeable missingness, more imputations may be required. Traditional literature on the topic stated that an imputation or m of around 5 is adequate (White, Royston and Wood, 2011). Whilst White et al. (2010) suggests using the Fraction of Missing Information (FMI) as a baseline for the minimum required imputations. If the maximum FMI in each model is 44 per cent, then 44 imputations are suggested at minimum. When following this assumption, White et al. (2010) found that standard errors and p-values were reduced and stabilised. Contemporary literature on the topic has gone back and forth on how many imputations is correct, Bodner (2008) has attempted to create a concrete set of procedures. He chose a key criterion that the width of the 95 per cent confidence interval should be within 10 per cent of its true value in 95 per cent of imputation runs. This led to the requirement of m imputations at 3, 6, 12, 24, 59 for FMI= 0.05, 0.1, 0.2, 0.3, 0.5, respectively (White, Royston and Wood, 2011). Bodner uses the FMI as a fraction of incomplete cases (Bodner, 2008). From this work and using tables 2 and 3 from his simulation study, there is a robust guideline of how many imputations to follow in each analysis. For example, in a sample whereby 30 per cent of data is missing, table 2 of Boder’s work (2008) shows that after 30 imputations marginal returns to efficiency are provided, table 3 confirms this for both the 95 per cent confidence intervals and the FMI statistic. For an extreme example, for a sample with 90 per cent missingness table 2 directs the researcher to perform >100 imputations and table 3 confirms 258 imputations is a baseline to achieve 95 per cent confidence interval half-widths. Bodner’s simulation tables will be used going forward to determine the required number of imputations alongside careful study of the FMI statistics post hoc.

After Multiple Imputation is performed, four key statistics are relevant to focus upon: variance total, Relative Variance Increase (RVI), Fraction of Missing Information (FMI), and Relative Efficiency (RE).

The primary usefulness of multiple imputation relies upon its variance estimation. The total variance in multiple imputation is the sum of multiple sources of variance: within imputation variance, between imputation variance and additional sampling variance. The latter is calculated by the within-imputation variance divided by the number of imputations. The variance total is directly related to how standard errors are calculated. Unlike simple imputation methods, multiple imputation estimates SEs so that the SEs for each parameter estimate are the square root of their variance totals.

The RVI or Relative Variance Increase is the proportional increase in total sampling variance due to missing information. Any variable that has a large amount of missingness or is weakly correlated with other variables in the imputation model tends to have larger than average RVIs. Weakly correlated auxiliary variables will always trend towards large RVIs.

The FMI is related to the RVI (which, in turn, is related to the variance total). The FMI is the proportion of the total sampling variance due to missing data. It is estimated based on the percentage of missingness for a particular variable and how correlated this variable is with other variables in the imputation model. When a variable has a high FMI, this can indicate a problematic variable, which may cause convergence issues.

Finally, the relative efficiency or RE relates to how well the actual population parameters are estimated. It is related to both the amount of missingness as well as the number of imputations within an imputation model. The RE is a comparative measure. It compares the relative efficiency of the current model variable to performing an infinite number of imputations. It is relatively easy to achieve a high RE on a given imputation model with few imputations; however, this does not mean that the standard errors within the given imputation model will be calculated accurately.

Paul Allison, in a series of articles (Allison, 2012a, 2012b, 2015), argues that FIML is 1) more straightforward to implement, 2) FIML has no incompatibility between an imputation model and an analysis model, 3) FIML produces a deterministic result rather than a different result every time, and 4) FIML is asymptomatically efficient.

Firstly, MI does have greater variability than FIML, but that increased choice in model selection is not necessarily a negative so long as proper procedures are followed. In fact, greater variability of choice has the potential to make MI a more attractive candidate for dealing with missingness over FIML. Secondly, MI models only run into an incompatibility problem when the MI model is inconsistent with the CRA model – something that, with appropriate testing and open science practices detailing the model construction, should not happen. Thirdly, MI models are deterministic, provided the same seed is used each time you run the imputation. The only time this would not be plausible would be when open science practices were not followed, and fellow researchers could not access the MI seed. Finally, the argument that FIML is asymptotically efficient only holds to a certain extent. MI models reach asymptotic efficiency by running an infinite number of imputations – though you can reach near full efficiency with a relatively small number of imputations, Allison (2015) argues, around 10. Overall, whilst FIML does offer some advantages, there is nothing so considerable theoretically as to desire FIML over MI on the condition that they both perform at near identical rates. So long as open science procedures are upheld, most major critiques of MI are dealt with.

There are very few comparisons between FIML and MI approaches to missing data. This makes it hard to assess if one method is more efficient at dealing with missing data than the other. Before conducting any missing data methods on the NCDS data a simulation is performed to assess the strengths of a range of handling missing data approaches with the intent to directly compare FIML and MI methods.

#### Simulation of Handling Missing Data Strategies

Both FIML and MI practices require data to either be MCAR or MAR. A FIML approach can be achieved in Stata by using the ‘sem’ command – using structural equation modelling and using the ‘mlvm’ estimation option (mlvm means FIML). MI can also be achieved in Stata using the ‘mi’ commands using a semi-Bayesian approach that includes auxiliary variables. There are also other handling missing data methods available such as: single mean imputation and coding all data=0 OR =1. These practices are typically considered ‘bad’ ways of handling missing data but are included in the simulation as a comparison to FIML and MI methods.

The full simulation takes the form of 1000 iterations of a random normal distribution of 1000 observations around a normally distributed metric dependent variable and three independent dummy variables that share an identical distribution. Each independent variable has the same level of correlation associated with the dependent variable. This is to allow for a point of comparison when MAR missingness is injected into one variable and not the others to see what happens when handling missing data practices are implemented. Each model is isolated in its own program whereby a simulation is called using the programs function with an identical seed set to all models. The 95 per cent confidence intervals of the mean betas and standard errors for all variables within each model are gathered and reported.

This dependent variable and three independent variables form a basic OLS linear regression model that is called the ‘Complete Records God Model’. Named as such because no model in a normal social scientific framework would have all observations not missing and have prior knowledge of what the ‘complete’ model would have looked like if their model did have some element of missingness. In addition to this ‘God’ model the same regression is computed using the structural equation modelling framework in Stata to confirm the results would be identical. The next model is where missingness is introduced. Missingness is injected into independent variable three. This missingness accounts for 49 per cent missingness in the model. This amount of missingness is right on the cusp of what contemporary literature on multiple imputation and FIML allow. Dummy variable adjustment is produced whereby all missingness is coded as =0 and another is produced =1. Next a single use modal imputation is used – the same framework as a single use mean imputation but because the variable is categorical mode is used over mean. Finally, an FIML model under the SEM framework is produced alongside three different forms of Multiple Imputation models. The first is an MI with 10 imputations and no auxiliary variables, the second is an MI with 10 imputations and auxiliary variables, and finally the last model is an MI with 100 imputations and auxiliary variables.

The results presented in table 2.45 display 10 separate models, each of which are illustrative of a handling missing data technique/method. The first two models, the ‘God Model’ and the ‘Complete SEM’ are not surprisingly identical. These two models use different commands in Stata to obtain the same results. With a MAR mechanism injected into model three, there is unsurprisingly a change in the estimates. There is an overall reduction in estimates by 0.09 across all three variables within the MAR injection model. Considering that all three independent variables were constructed using the same uniform distribution, it is expected that an injection of MAR missingness at one variable would correspond to identical distortions in all estimates.

The fourth and fifth models demonstrate the limited utility of dummy variable adjustment models. For the fourth model, where all missingness is coded as equal to zero, it is demonstrated how dangerous this method can potentially be, the estimates of independent variables one and two are overinflated and the estimate of independent variable 3 is flipped and reduced. The fifth model, whereby all missingness is equal to one does obtain identical results to the ‘God Model’. However, it would be naïve to state that this provides utility in of itself. Without knowing a priori information about missingness in each model we cannot ever know with certainty that a dummy adjustment model would obtain accurate estimates. In the case of a simplistic setting, whereby all missingness is attached to a single variable, and that variable happens to be a dummy variable, we can know with some level of certainty of the estimates that are obtained. In a similar scenario without a priori knowledge, dummy variable adjustment does have utility in providing a range of values for estimates. In this example, if a priori knowledge did not exist, the researcher could state that the estimate coefficient of independent variable three exists within a range of 0.07 and -0.19. Whilst this is not a perfect solution, it does provide more information than not pursuing handling missing data strategies.

The single use modal imputation demonstrates its lack of utility in of itself. Due to the dummy-based nature of independent variable three, this naïve model produces estimates identical to model four. At least with a combined dummy variable adjustment approach a range of estimates can be stated, with a single use imputation approach no utility can be ascertained.

The FIML approach does regain some of the original models estimates but does appear to struggle at the level of 50 per cent missingness. FIML appears to be best suited for retrieving the original estimates from the variable where missingness is located. In this example, independent variable three is the variable that FIML appears to best retrieve the ‘real’ estimates from. The other two variables of interest in the model, even though they share the same uniform distribution as independent variable three, do not receive the same adjustment as independent variable three. FIML does a good job and the substantive interpretation, whilst altered somewhat fundamentally stays the same.

Finally with respect to the three Multiple Imputation models, each provides interesting discussion. Firstly, the MI model with zero auxiliary variables appears to provide near identical estimates to that of the ‘God Model’. This at first is somewhat surprising considering the lack of auxiliary variables but considering the relative simplicity of the model being simulated the relative precision of a MI model with no auxiliary variables does appear to be strong. The Imputation model with auxiliary variables at 10 imputations is demonstrated to be the most precise handling missing data method out of all produced in this simulation. This model produces identical estimates to the ‘God Model’. Finally, the Imputation model with 100 imputations produces slightly worse estimates than its previous iteration. This is a surprising result, considering there is no such thing as ‘over-estimation’ within multiple imputation. This could simply be an artefact of the simulation study and its respective seed. In saying that, the difference in reported estimates is as small as 0.01. There is truly little evidence to support the view that this model performs ‘worse’ than an imputation with 10 iterations.

Table 2.46 Simulation Regression Models Using a MAR Principle

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Table 2.45 Simulation Regression Models Using a MAR Principle | | | | | | | | | | |
|  | Complete Records 'God Model' | Complete SEM | Missingness Introduced at Independent Variable 3 | All Missingness coded as =0 | All Missingness coded as =1 | Single Use Modal Imputation | FIML | Imputed with no auxiliary variables and 10 imputations | **Imputed with 10 imputations** | Imputed with 100 imputations |
| Independent Variable 1 | [-0.19, -0.19] | [-0.19, -0.19] | [-0.10, -0,10] | [-0.28, -0.27] | [-0.19,-0.19] | [-0.28, -0.27] | [-0.12,-0.12] | [-0.20, -0.20] | **[-0.19, -0.18]** | [-0.20, -0.20] |
|  | [(0.02,0.02)] | [(0.02,0.02)] | [(0.01,0.01)] | [(0.02,0.02)] | [(0.02,0.02)] | [(0.02,0.02)] | [(0.02,0.02)] | [(0.02,0.02)] | **[(0.02,0.02)]** | [(0.02,0.02)] |
| Independent Variable 2 | [-0.19, -0.19] | [-0.19, -0.19] | [-0.10, -0,10] | [-0.28, -0.28] | [-0.19,-0.19] | [-0.28, -0.28] | [-0.12, -0.12] | [-0.18, -0.18] | **[-0.19, -0.19]** | [-0.19, -0.19] |
|  | [(0.02,0.02)] | [(0.02,0.02)] | [(0.01,0.01)] | [(0.02,0.02)] | [(0.02,0.02)] | [(0.02,0.02)] | [(0.02,0.02)] | [(0.02,0.02)] | **[(0.02,0.02)]** | [(0.02,0.02)] |
| Independent Variable 3 | [-0.19, -0.19] | [-0.19, -0.19] | [-0.10, -0,10] | [0.07,0.07] | [-0.19,-0.19] | [0.07,0.07] | [-0.25, -0.25] | [-0.20, -0.20] | **[-0.19, -0.19]** | [-0.18, -0.18] |
|  | [(0.02,0.02)] | [(0.02,0.02)] | [(0.01,0.01)] | [(0.02,0.02)] | [(0.02,0.02)] | [(0.02,0.02)] | [(0.01.0.01)] | [(0.02,0.02)] | **[(0.02,0.02)]** | [(0.02,0.02)] |
| Number of observations | 1000 | 1000 | 513 | 1000 | 1000 | 1000 | 1000 | 1000 | **1000** | 1000 |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: Simulation using a MAR principle. 51 per cent missingness introduced. | | | | | | | | | | |

This simulation has attempted to accomplish two tasks. The first, related to the establishment of handling missing data methods into categorisations of ‘standards’ of approaches. The second, related to a direct comparison to the two ‘gold standard’ methods of handling missing data, FIML and MI. The first task has been accomplished by detailing each handling missing data procedure and analysing their effectiveness in a simulation study using 1000 observations in a simulation that repeats the model 1000 times. Results demonstrate that when a MAR mechanism is present, a complete records-based approach is insufficient and potentially damaging to the substantive interpretation of results – as is an ad hoc method such as single use imputation. Other handling missing data methods such as dummy variable adjustment does have potential drawbacks, but also have some utility given that the researcher understands exactly what dummy variable adjustment is doing and that it is not a replacement for ‘gold standard’ approaches. Finally, the ‘gold standard’ methods are re-affirmed to be the best adjustment procedures for handling missing data within data analysis. With respect to the second task, this paper, has discussed the theoretical and practical advantages and disadvantages of FIML and MI as has been laid out in other papers but has also gone further through the simulation analysis, by providing a direct comparison of FIML and MI. The results demonstrate that MI provides marginally better returns on precise estimates even if the correct MI procedures are not followed. Some caution with this conclusion is needed. The small size of these estimates and the differences between them means that the only strong conclusion that can be drawn from this study is that both FIML and MI offer practical, efficient methods to handle missing data and the choice of procedure should depend on the data and models that the researcher is planning to conduct. For simple linear based models, FIML offers a quick and efficient solution of MAR based mechanisms. For more complex models, MI offers a larger variety of options to choose from without having to break user flow by moving to different statistical software.

#### Handling Missing Data in the NCDS

There are 12,450 individuals identified in the NCDS who indicated some form of economic activity with the sweep at age 23. After using the variable related to the outcome of tracing and interviews, there are 12,536 individuals within this sample[[18]](#footnote-18). There are 4,638 observations with missing data on at least one of the variables included for analysis. Of the missingness amongst variables, 86 were missing in economic activity, 26 in educational attainment, 1893 in housing tenure and 3779 on NS-SEC. Sex has no missing data as it was recorded at wave 0 (so all individuals were included).

Patterns of missing data are presented in Table 2.46. Within the NCDS sample, 67 per cent have complete records on all variables, 17 per cent are missing values at socio-economic measures, a further 13 per cent are missing on socio-economic measures and housing tenure, and 2 per cent are missing at housing tenure. Further missingness in the sample not presented in the table is <1 per cent. In total, 8,411 cases have a complete observation of all variables.

An overview by (Power and Elliott, 2006) suggests that after accounting for death and emigration, sample loss over time is mainly attributed to individuals moving within the UK and not responding to requests to trace them. As mentioned previously, refusal at age 23 was 7.1 per cent. With an eligible sample of 16,402, this corresponds to 1,181 people dropping out due to refusal. Taking the eligible sample after death, emigration, and refusal to 15,221. At age 23, the NCDS notes that 12,503 people were successfully traced and conducted a full interview, with a further 33 completing a partial interview (Power and Elliott, 2006). This amounts to 12,536 people successfully interviewed, meaning that 2,686 people who did not die, emigrate, or refuse to participate in the survey are missing from Sweep 4 (age 23).

Table 2.47 Missing data patterns for NCDS

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N | Percent Complete (%) | Educational Attainment | Economic Activity | Housing Tenure | NS-SEC | Sex |
| 8411 | 67 | **✓** | **✓** | **✓** | **✓** | **✓** |
| 2201 | 17 | **✓** | **✓** | **✓** |  | **✓** |
| 1636 | 13 | **✓** | **✓** |  |  | **✓** |
| 251 | 2 | **✓** | **✓** |  | **✓** | **✓** |
| Total = 12536 |

Of the missing data, 67 per cent of cases are complete, and the largest proportion of missing data comes from the three socio-economic variables: NS-SEC, CAMSIS, and RGSC. Individuals who are inactive in the labour force find it challenging to code within socio-economic measures. NS-SEC categorisation is based on occupation type, authority duties, and the organisation's size (Rose and Pevalin, 2002). CAMSIS occupational scores only relate to the working population within a country. The RGSC also suffers similar problems concerning categorisation of retired, unemployed, and out of the labour force individuals. Unemployment rates during this period, including people out of the labour force, are around 5-7 per cent (ONS, 2023). This, combined with refusal rates of 7.1 per cent (Power and Elliott, 2006), presents a reasonable case in explaining the potential missingness of the largest missing variables in the model.

A complete records analysis is only valid if data can be considered MCAR. If data is found to be MAR, then steps should be taken to address this potential bias in analysis. Following the prior simulation study, multiple imputation will be used for handling missing data purposes going forward. The following section seeks to compare a CRA and MI approach to estimate if there are any differences in the substantive conclusions reached. If differences are found, implications are then discussed.

When selecting auxiliary variables, disadvantaged socio-economic background in childhood, worse mental health and lower cognitive ability in early life, and lack of civic and social participation in adulthood are consistently associated with non-response (Silverwood *et al.*, 2021). These variables are easily translated into auxiliary variables for imputation. Using the NCDS missing data guide (ibid), each predictor of non-response at sweep 4 (age 23) was recoded. These correspond to region, number of persons per room, sex of the child, social class of mother’s husband, family moves since child’s birth, dad reads to a child, area of world in which mother was born, number of family moves since child’s birth, number of household amenities, number of family moves since child’s birth, sum of favourable learning environments.

The imputation model naturally includes all analytical variables included in the previous section. The imputation model also includes several auxiliary variables to add to maximising the plausibility of the MAR assumption to reduce bias due to missingness (ibid). These are broken down into two types. The first are variables that are predictive of both the probability of missingness and the underlying missing values themselves. The second are variables that are predictive of the underlying missing values only. Missing data in the NCDS is derived mainly by non-response at a given sweep, and auxiliary variables are selected from pre-determined sets of variables predictive of non-response at sweep 4, as seen in (Silverwood *et al.*, 2021). The fact that missingness within the NCDS is primarily driven by sweep non-response rather than item non-response means it is even more critical to include auxiliary variables within the imputation.

There were 18 variables identified for inclusion in the imputation model. However, variables that are predictive of the chance of missing values but are not predictive of the underlying missing values themselves will not add any information to the model (ibid). Thus, the decision was made not to include such variables in the imputation model that were not predictive of economic activity following the advice from the NCDS guide on handling missing data (ibid). From this, 8 out of 18 variables are substantively associated with economic activity and are included as auxiliary variables in the imputation model.

Prior to imputation, it is best to explore the distribution of variables compared to complete and non-complete cases. In the presence of an MCAR mechanism, all distributions should be the same comparatively. If this is not the case, then this is suggestive of a MAR or MNAR mechanism. These imbalances present themselves in every variable within the model except for sex. This is unsurprising, considering that sex as a variable presents zero missingness. The distributions of the variables thus far present some indications of a MAR or MNAR mechanism.

With all the variables in the model being categorical, convergence issues are a possibility. This risk is increased if a model has many categorical variables. Failure to converge was a consistent problem. Without resorting to re-coding analytical variables, the decision was made to drop one of the auxiliary variables to produce an imputed model[[19]](#footnote-19).

After performing the imputation, it is often helpful to graph the means and standard deviations saved through the tracing subcommand when using MICE – autocorrelation plots would be helpful but are only available for non-MICE related imputations. By graphing variables means and standard deviations through trace plots, for example, over each imputation, any discrepancy or deviation can easily be found. If this were to be the case, this would be problematic for the imputation model and suggest that further imputations would be required (White, Royston and Wood, 2011). The means and standard deviations of imputed values from each iteration[[20]](#footnote-20) were checked to see the distributions of each variable against the imputations[[21]](#footnote-21).

A table of descriptive statistics comparing the complete records sample and the imputed sample can be found in table 2.47. The following models presented will compare a complete records analysis using NS-SEC from the previous chapter and the imputed model in Table 2.48 The CRA model has 8,411 observations. Using a variable within the NCDS dataset [n4118] that noted how many individuals were successfully contacted for sweep 4 (age 23) of the NCDS, there are 12,536 individuals within this sweep. The imputed dataset thus has 12,536 observations compared to the 8,411 observations of the CRA model. Following Bodner’s views on number of imputations (Bodner, 2008), a dataset such as this with 33 per cent missingness around 30 imputations reaches peak efficiency, 24 imputations are required to achieve 95 per cent Cis half-widths and 36 are required for 95 per cent fractions of missing information to achieve specified precision. Accordingly, using the maximum number of imputations required and rounding to the nearest 10 requires the NCDS dataset to have 40 imputations.

The results for both the complete records analysis and the imputed model can be viewed in Table 2.48. Overall, there is a similarity between the complete records analysis and the imputed model. The substantive conclusions between CRA and MI models are nearly identical. There are some very slight differences in the log odds and average marginal effects across the variables. However, these slight differences are not large enough to impact the substantive conclusions presented in the interpretation of the CRA model. The imputed model confirms the substantive conclusions made from the CRA model with some minor variation in log odds and average marginal effects and a reduction in standard errors. The results demonstrate substantively identical findings from both the CRA and MI models. This provides a solid justification for the missingness within the NCDS model to be MCAR rather than MAR. This provides a level of confidence in the substantive findings of the CRA model going forward. Any comparisons made going forward using the NCDS model will refer to the CRA model rather than the MI model. In addition, using the NCDS sample going forward, unless using a different model, will not require multiple imputation. There is confidence that the complete records analysis presents the most appropriate substantive interpretation of the model. The level of missingness present within the model and at individual variables within the model does not seem to have a substantial enough impact upon model interpretation when compared with an imputed model. The interpretation and findings made prior to imputation stand.

To aid interpretation further, figure 2.29 presents a direct comparison of the coefficients of both the CRA and MI models against one another. As can be seen visually the coefficients for all MI variables is within the confidence intervals of the CRA variables. Average marginal effects and predictive margins are also plotted comparatively in figure 2.30.

Table 2.48 Descriptive statistics comparing CRA versus MI models

|  |  |  |
| --- | --- | --- |
| Table 2.47: Descriptive Statistics for Economic Activity | | |
|  | CRA% | MI% |
| Continue Schooling or not after September when individuals are 16 |  |  |
| Don't Continue Schooling | 60.83% | 61.72% |
| Continue Schooling | 39.17% | 38.28% |
| Educational Attainment O-levels |  |  |
| <5 O-Levels | 64.51% | 66.17% |
| >5 O-Levels | 35.49% | 33.83% |
| Sex of Respondent |  |  |
| Female | 50.11% | 50.02% |
| Male | 49.89% | 49.98% |
| Housing Tenure of Respondent when Child |  |  |
| Own Home | 48.09% | 46.84% |
| Don't Own Home | 51.91% | 53.16% |
| NS-SEC Social Class of Father when Respondent Child SOC2000 |  |  |
| Large Employers and higher managerial occupations | 3.10% | 3.20% |
| Higher professional occupations | 4.87% | 6.67% |
| Lower Managerial and professional occupations | 12.34% | 11.33% |
| Intermediate occupations | 9.57% | 10.23% |
| Small employers and own account workers | 12.17% | 13.24% |
| Lower supervisory and technical occupations | 16.31% | 16.63% |
| Semi-routine occupations | 17.66% | 19.66% |
| Routine occupations | 23.97% | 19.03% |
|  |  |  |
| n | 8411 | 12536 |
| Data Source: NCDS | | |

A graph with red and blue lines

Description automatically generated

Figure ‎2.28 Coefficient plot comparing CRA and MI models

A white background with black and white lines

Description automatically generated

Figure ‎2.29 Margins plot comparing CRA and MI models

Table 2.49 MI versus CRA for NCDS model

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | NS-SEC (SOC200) CRA | | | Average Marginal Effects | | NS-SEC (SOC 2000) MI | | | Average Marginal Effects | |
| Economic Activity: ‘Don’t Continue Schooling’ Reference Category | Coef. | S.E. | Sig. | Prob. | S.E. | Coef. | S.E. | Sig. | Prob. | S.E. |
| Continue Schooling |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| Less than five O’levels | Ref. |  |  |  |  |  |  |  |  |  |
| Five or More O’levels | 2.98 | (0.07) | \*\*\* | 0.56 | (0.01) | 2.98 | (0.05) | \*\*\* | 0.59 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| Female | Ref. |  |  |  |  |  |  |  |  |  |
| Male | -0.50 | (0.06) | \*\*\* | -0.06 | (0.01) | -0.45 | (0.05) | \*\*\* | -0.06 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| Own Home | Ref. |  |  |  |  |  |  |  |  |  |
| Do not Own Home | -0.63 | (0.06) | \*\*\* | -0.08 | (0.01) | -0.63 | (0.06) | \*\*\* | -0.09 | (0.01) |
| NS-SEC |  |  |  |  |  |  |  |  |  |  |
| 1.1 | 0.02 | (0.19) |  | 0.00 | (0.03) | 0.07 | (0.18) |  | 0.01 | (0.03) |
| 1.2 | 0.48 | (0.17) | \*\*\* | 0.08 | (0.03) | 0.54 | (0.16) | \*\*\* | 0.09 | (0.03) |
| 2 | Ref. |  |  |  |  |  |  |  |  |  |
| 3 | -0.25 | (0.13) | \* | -0.04 | (0.02) | -0.23 | (0.12) |  | -0.04 | (0.02) |
| 4 | -0.89 | (0.12) | \*\*\* | -0.13 | (0.02) | -0.85 | (0.12) | \*\*\* | -0.12 | (0.02) |
| 5 | -0.76 | (0.11) | \*\*\* | -0.11 | (0.02) | -0.74 | (0.10) | \*\*\* | -0.11 | (0.02) |
| 6 | -0.89 | (0.11) | \*\*\* | -0.13 | (0.02) | -0.88 | (0.11) | \*\*\* | -0.13 | (0.02) |
| 7 | -1.11 | (0.11) | \*\*\* | -0.16 | (0.02) | -1.12 | (0.11) | \*\*\* | -0.16 | (0.02) |
| Intercept | -0.40 | (0.09) | \*\*\* |  |  | -0.42 | (0.09) | \*\*\* |  |  |
| Number of observations | 8411 | | | | | 12536 | | | | |
| Average RVI |  | | | | | 0.26 | | | | |
| Largest FMI |  | | | | | 0.34 | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: NCDS [Sweeps 0-4]  Note: CRA versus MI model for NCDS | | | | | | | | | | |

#### Discussion and Conclusions

This section had two aims. The first was to investigate the optimal handling missing data strategy by producing a simulation of handling missing data methods. The second was to put the most optimal handling missing data method into practice and compare it with a complete records analysis of the NCDS data to assess the evidence for a MAR assumption. Both aims were successful. A simulation was conducted and found re-affirmed that FIML and MI based methods for handling missing data were the most optimal strategies out of all assessed. The conclusions also found that with the data being used in this analysis and the simulation results, multiple imputation would be most optimal going forward.

Results from the comparison of CRA and MI models show that there are no substantive differences between the complete records analysis and the imputed model suggests little evidence for a MAR assumption. This indicates that missingness in these categories has no substantive impact on the resulting interpretation and analysis of results. As such, the imputed model is no better than analysing the complete record for interpretation. Whilst this section does present much work that amounts to a preference for a complete records analysis model, the use of multiple imputation and discussion of missing data was essential to discuss. There was no a priori evidence to suggest that the missing data present within this model was not MAR. Due to this, and due to the ability to check this assumption, there was statistical due diligence to present these findings.

Implementing tools seeking to deal with missingness within this model provides contemporary statistical techniques to the research of youth transitions within the NCDS. Whilst substantively finding identical results from complete records analysis to multiple imputed models, this exercise in dealing with missingness has allowed for the updating of youth transitions literature within this field of study. There is a relative level of confidence that this research looking into the choices and opportunities of youth transitions from mandatory education has attempted to control for all statistical possibilities.

### Discussion and Conclusions for Granular NCDS Analysis

The overall empirical finding from the analysis is that structural inequalities matter for NCDS youth in influencing their choice and opportunity as it relates to youths first transition. Social class, sex, housing tenure, and educational attainment all have substantively significant impacts on an individual’s choice and opportunity concerning pathway selection post-mandatory education.

The fact that previous structural inequalities that have manifested during previous life domains (Mayer, 2009) influence life chances in other life domains indicates support for promoting a life course perspective within this research.

The findings from this empirical research appear to confirm the influential impact that structures have on individual life chances. Individual choice is constrained and influenced by structural factors such as social class, sex, and housing tenure. Something that is supported by previous literature (Hutchison, Prosser and Wedge, 1979; Connolly, Micklewright and Nickell, 1992; Booth and Satchell, 1994; Schoon *et al.*, 2001; Dolton, Joshi and Makepeace, 2002; Dolton, Makepeace and Marcenaro‐Gutierrez, 2005). Echoing back to discussions on agency and structure, this empirical evidence is suggestive of an agency within structure understanding of the life course – individuals certainly choose their given pathways and opportunities. However, structures continue to influence and constrict these opportunities. Furthermore, whilst these empirical results confirm much of previous empirical literature on this subject, the arguments proposed by new structuralist theorists (Saunders, 2003, 2021) are not as sound. Evidence has been found that concurs with the premise first emphasised by Saunders (2003) that housing tenure offers an independent and substantive impact on life chances. The argument that it is dominant above social class has no evidence within these results.

Including prior educational attainment alongside structural effects such as social class, sex, and housing tenure provides a much more complex relationship with individuals' choices and opportunities post-mandatory education. The findings provide substantial empirical evidence that prior educational attainment does matter concerning pathway selection.

This work provides methodological innovation by conducting a sensitivity analysis of socioeconomic measures of social stratification. Sensitivity analysis of NS-SEC, CAMSIS, and RGSC demonstrate that NS-SEC is a robust and strong measure of social class, which is suitable for use within this model using NCDS data. The findings of this sensitivity analysis provided slightly divergent substantive findings. Choosing NS-SEC as the dominant model through the analysis was based upon a theoretical desire to understand class-based dynamics. Through its implementation, social class was found to have a resounding impact on individuals’ choices and opportunities concerning transitional experiences. Whilst social class measures provided remarkably similar substantive findings, the theoretical differences between NS-SEC and RGSC measures primarily surrounding the manual/non-manual divide did produce different lens in which to view said results. CAMSIS offered a direct contract to social class-based methods of analysing social stratification, as a social distance scale it provided a remarkably different substantive interpretation compared to social class-based measures.

On top of this, another sensitivity analysis was conducted to reflect on the similarities and differences between different constructions of social stratification measures based upon SOC – using both SOC 2000 and SOC 90 to compare statistical and substantive results. Findings demonstrate that SOC 2000 is preferred statistically, and whilst both models agree on the general trend of substantive results, there is disagreement in the size of these trends and effect sizes. In addition, results suggest that social class-based measures of social stratification such as NS-SEC and RGSC are much more sensitive to SOC constructions compared to measures of social distance such as CAMSIS. This serves as implications for variable inclusion in models that look at social change over long periods of time that cross different SOC boundaries.

The results are also innovative by assessing handling missing data methods first through a simulated study of 10 different methods then through an implementation of the most optimal method compared to a complete records analysis of NCDS data. Missingness was first descriptively detailed before strategies for handling such missingness were discussed. A multiple imputation model found that missingness has no impact on the substantive findings of the complete records analysis model. While this means that the substantive findings remain the same as previously detailed, the implementation of dealing with missing data was an essential contemporary statistical strategy that previous literature within this field typically overlooked. Both the implementation of sensitivity analysis and multiple imputation techniques thus serve as methodological innovations beyond prior literature within the field.

Overall, the literature has been updated, with prior literature being confirmed in some cases (Connolly, Micklewright and Nickell, 1992; Booth and Satchell, 1994; Dolton, Makepeace and Marcenaro‐Gutierrez, 2005) and challenged in others (Saunders, 2003, 2021). Social theories related to youth transitions and the impacts on structural effects and agency within the realm of choice and opportunity have also been contended. Finally, modern statistical techniques have been applied to update prior literature to modern statistical standards by employing sensitivity analyses and providing tools to deal with potential missingness.

Whilst the theory of individualisation provides an interesting theoretical explanation, it is incorrect to claim that a risk society is a classless society or indeed a sexless society (Furlong and Cartmel, 2007). The old social cleavages of social class and sex remain intact though this may become complicated as more contemporary youth cohorts are examined.

Going forward, whilst previous literature has been confirmed and updated, questions remain essential to reflect on for future inquiry. As has been mentioned, structural influence is dependent upon the given pathway of choice, with different structural influences matter more for some pathways than for others. A closer inspection of these differences is called for. On top of this, so far, this research reflects upon the entry into or the transitional stage of going from school to work. While the identification of several pathways has been made apparent, prior literature related to the relative smoothness of these transitions is paramount to focus on next. Structural effects matter for the transition itself, but whether they manifest throughout the life domain (Mayer, 2009) is relatively significant in the discussion of structural effects throughout the life course.

The following section will attempt to duplicate the analysis conducted in this section using the British Cohort Study (BCS). The BCS is a nationally representative birth cohort survey conducted in a week in 1970. Much of the data in the BCS has been harmonised with the NCDS, which allows for a detailed comparison of trends between cohorts. The analysis of the next section will reflect on the differences and similarities that have been made within the conclusion of this section, focusing once again on the nature of structural inequalities and their influence on choice and opportunities for youth transitions.

## In-Depth BCS Analysis

The in-depth NCDS analysis demonstrated some unique findings with regards to social stratification, SOC codes, and missingness. This in-depth analysis is now repeated using the BCS sample. Using a different sample and employing the same methods allows a comparison of findings between samples and cohorts. The findings from this in-depth BCS analysis provides an opportunity to assess the in-depth findings from the NCDS sample to understand if these findings are repeatable across samples or unique to the socio-historical context of the NCDS.

This section will focus on the British Cohort Study (BCS). An initial logistic regression of the BCS sample will be provided and interpreted. Following this, a sensitivity analysis of social stratification measures will be provided as well as an analysis using alternative standard occupation codes. Finally, a section on handling missing data will be provided. This in-depth analysis of the BCS sample will be concluded with a discussion of the relevant findings across the analysis and re-iterate the main findings.

### Descriptive Statistics

Table 2.39 shows the frequencies and summary statistics for the BCS. The overall N for the BCS sample is 1574. Overall, 60.36 per cent of the sample continues schooling compared to 39.64 per cent that don’t continue schooling. Comparing these statistics with table 2.15 illustrates a near complete reversal of individuals continuing schooling compared to the NCDS cohort.

Regarding educational attainment 61.05 per cent of individuals received less than five O’levels compared to 38.95 per cent of their peers that did receive five or more O'levels. Sex presents an even split of 55.40 per cent of women and 44.60 per cent of men. Regarding homeownership, 76.24 per cent of individuals grew up in homes owned by their parents compared to 23.76 per cent of individuals that did not grow up in homes owned by their parents. Comparing this to information provided by the Home Owners Alliance (2012) demonstrates that this sample has an overrepresentation of individuals that come from homes owned by their parents. The rate reported in 1986 for home ownership rates in England and Wales is 65.5 per cent – this sample has an overrepresentation of 10.74 per cent. This could present missing data related issues within model interpretation and will be investigated later in this chapter.

The NS-SEC categories for SOC 2000 construction all see a relatively even distribution between 10-19 per cent except for those in NS-SEC 1.1 and 1.2 with 5.65 per cent and 7.50 per cent respectively. The SOC 90 construction echoes this trend alongside having NS-SEC 2 as the single largest category of NS-SEC. There are some small variations between SOC 2000 and SOC 90 constructions of RGSC, but both follow a similar pattern. RGSC 5 is the single smallest category across SOC constructions and RGSC 3M is the single largest category. SOC 2000 and SOC 90 constructions of CAMSIS are near identical for their means and standard deviations though the SOC 90 construction has a slightly increased standard deviation compared to SOC 2000.

Table 2.50 Descriptive Statistics for BCS Model

|  |  |  |
| --- | --- | --- |
| Table 2.39: Descriptive Statistics for Economic Activity | | |
|  | n | % |
| Continue Schooling or not after September when individuals are 16 |  |  |
| Don't Continue Schooling | 624 | 39.64% |
| Continue Schooling | 950 | 60.36% |
| Educational Attainment O'levels |  |  |
| <5 O-Levels | 961 | 61.05% |
| >5 O-Levels | 613 | 38.95% |
| Sex of Respondent |  |  |
| Female | 872 | 55.40% |
| Male | 702 | 44.60% |
| Housing Tenure of Respondent when Child |  |  |
| Own Home | 1,200 | 76.24% |
| Don't Own Home | 374 | 23.76% |
| NS-SEC Social Class of Father when Respondent Child SOC2000 |  |  |
| 1.1 | 89 | 5.65% |
| 1.2 | 118 | 7.50% |
| 2 | 296 | 18.81% |
| 3 | 179 | 11.37% |
| 4 | 170 | 10.80% |
| 5 | 258 | 16.39% |
| 6 | 198 | 12.58% |
| 7 | 266 | 16.90% |
| RGSC Social Class of Father when Respondent Child SOC2000 |  |  |
| 1 | 92 | 5.84% |
| 2 | 458 | 29.10% |
| 3NM | 196 | 12.45% |
| 3M | 577 | 36.66% |
| 4 | 174 | 11.05% |
| 5 | 77 | 4.89% |
| NS-SEC Social Class of Father when Respondent Child SOC90 |  |  |
| 1.1 | 85 | 5.40% |
| 1.2 | 132 | 8.39% |
| 2 | 315 | 20.01% |
| 3 | 178 | 11.31% |
| 4 | 157 | 9.97% |
| 5 | 261 | 16.58% |
| 6 | 199 | 12.64% |
| 7 | 247 | 15.69% |
| RGSC Social Class of Father when Respondent Child SOC90 |  |  |
| 1 | 123 | 7.81% |
| 2 | 391 | 24.84% |
| 3NM | 266 | 16.90% |
| 3M | 536 | 34.05% |
| 4 | 204 | 12.96% |
| 5 | 54 | 3.43% |
|  |  |  |
|  | Mean | SD |
| CAMSIS SOC2000 | 49.06 | 13.81 |
| CAMSIS SOC90 | 49.38 | 14.52 |
|  |  |  |
| n |  | 1574 |
| Data Source: BCS | | |

Table 2.40 provides descriptive statistics by the dependent variable. From table 2.40 some observations can be made. Whilst not continuing schooling appears to be stratified by educational attainment, with 77.88 per cent of individuals not receiving five or more O’levels not continuing schooling, and 22.12 per cent of individuals that did receive five or more O’levels did not continue schooling the same stratification effect is not evident for individuals that did continue schooling. Individuals that received five or more O’levels had the same per cent continuing schooling compared to those that did not receive five or more O’levels.

The reverse is true looking at the stratification of sex by economic activity. Whilst there is no underlying difference between men and women not continuing schooling, there is a difference for men and women continuing schooling. For those that continued schooling, 60.32 per cent of the sample were women compared to 39.68 per cent of men.

The total sample has 76.24 per cent of individuals living in homes owned by their parents. Looking now at table 2.41 there is an underrepresentation of homeowners who didn’t continue schooling at 71.15 per cent and an overrepresentation who continued schooling at 79.58 per cent.

Looking at the NS-SEC SOC 2000 construction, a majority of the sample located within the don’t continue schooling category are situated within NS-SEC 5-7. Those that continue schooling see a larger makeup of children from social origins of NS-SEC 1.1,1.2, and 2. The same is true for the SOC 90 construction of NS-SEC. For RGSC SOC 2000 skilled non-manual workers nearly makes up most individuals that did not continue schooling. This is diminished for the SOC 90 construction of RGSC but across both measures it still maintains the plurality of individuals not continuing schooling. The single largest grouping that continued schooling came from RGSC 2 – this is echoed across SOC 2000 and SOC 90 measures. For CAMSIS there is a near identical difference in means of individuals that didn’t continue schooling compared to those that did – a 5-point difference.

Table 2.51 Descriptive Statistics for BCS Model by dependent variable

|  |  |  |  |
| --- | --- | --- | --- |
| Descriptive Statistics by Economic Activity | | | |
|  | Continue Schooling or not after September when individuals are 16 | | |
|  | Don't Continue Schooling | Continue Schooling | Total |
| N | 624 (39.64%) | 950 (60.36%) | 1574 (100.00%) |
| Educational Attainment O'levels |  |  |  |
| <5 O-Levels | 486 (77.88%) | 475 (50.00%) | 961 (61.05%) |
| >5 O-Levels | 138 (22.12%) | 475 (50.00%) | 613 (38.95%) |
| Sex of Respondent |  |  |  |
| Female | 299 (47.92%) | 573 (60.32%) | 872 (55.40%) |
| Male | 325 (52.08%) | 377 (39.68%) | 702 (44.60%) |
| Housing Tenure of Respondent when Child |  |  |  |
| Own Home | 444 (71.15%) | 756 (79.58%) | 1200 (76.24%) |
| Don't Own Home | 180 (28.85%) | 194 (20.42%) | 374 (23.76%) |
| NS-SEC Social Class of Father when Respondent Child SOC2000 |  |  |  |
| 1.1 | 23 (3.69%) | 66 (6.95%) | 89 (5.65%) |
| 1.2 | 30 (4.81%) | 88 (9.26%) | 118 (7.50%) |
| 2 | 87 (13.94%) | 209 (22.00%) | 296 (18.81%) |
| 3 | 64 (10.26%) | 115 (12.11%) | 179 (11.37%) |
| 4 | 80 (12.82%) | 90 (9.47%) | 170 (10.80%) |
| 5 | 125 (20.03%) | 133 (14.00%) | 258 (16.39%) |
| 6 | 86 (13.78%) | 112 (11.79%) | 198 (12.58%) |
| 7 | 129 (20.67%) | 137 (14.42%) | 266 (16.90%) |
| RGSC Social Class of Father when Respondent Child SOC2000 |  |  |  |
| 1 | 24 (3.85%) | 68 (7.16%) | 92 (5.84%) |
| 2 | 132 (21.15%) | 326 (34.32%) | 458 (29.10%) |
| 3NM | 68 (10.90%) | 128 (13.47%) | 196 (12.45%) |
| 3M | 283 (45.35%) | 294 (30.95%) | 577 (36.66%) |
| 4 | 79 (12.66%) | 95 (10.00%) | 174 (11.05%) |
| 5 | 38 (6.09%) | 39 (4.11%) | 77 (4.89%) |
| NS-SEC Social Class of Father when Respondent Child SOC90 |  |  |  |
| 1.1 | 20 (3.21%) | 65 (6.84%) | 85 (5.40%) |
| 1.2 | 38 (6.09%) | 94 (9.89%) | 132 (8.39%) |
| 2 | 91 (14.58%) | 224 (23.58%) | 315 (20.01%) |
| 3 | 64 (10.26%) | 114 (12.00%) | 178 (11.31%) |
| 4 | 79 (12.66%) | 78 (8.21%) | 157 (9.97%) |
| 5 | 127 (20.35%) | 134 (14.11%) | 261 (16.58%) |
| 6 | 90 (14.42%) | 109 (11.47%) | 199 (12.64%) |
| 7 | 115 (18.43%) | 132 (13.89%) | 247 (15.69%) |
| RGSC Social Class of Father when Respondent Child SOC90 |  |  |  |
| 1 | 35 (5.61%) | 88 (9.26%) | 123 (7.81%) |
| 2 | 112 (17.95%) | 279 (29.37%) | 391 (24.84%) |
| 3NM | 92 (14.74%) | 174 (18.32%) | 266 (16.90%) |
| 3M | 259 (41.51%) | 277 (29.16%) | 536 (34.05%) |
| 4 | 97 (15.54%) | 107 (11.26%) | 204 (12.96%) |
| 5 | 29 (4.65%) | 25 (2.63%) | 54 (3.43%) |
| CAMSIS SOC2000 | 45.78 (12.51) | 51.21 (14.21) | 49.06 (13.81) |
| CAMSIS SOC90 | 46.04 (13.09) | 51.57 (15.00) | 49.38 (14.52) |

Looking in further detail on the analytical construction of each of the three social stratification variables, a cross-tabulation is created for both NS-SEC and RGSC measures and a summary statistics table is provided for CAMSIS. Table 2.41 provides a cross-tabulation of NS-SEC by its SOC 2000 and SOC 90 constructions. Table 2.41 details a cross-tabulation of NS-SEC by SOC construction. Looking at the diagonals demonstrates how many observations share the same NS-SEC category for both the SOC 2000 and SOC 90 constructions. No diagonal has less than 78 per cent overlap. The lowest overlap occurs in NS-SEC 2 with 79.68 per cent. The largest single overlap occurs at NS-SEC 7 with 95.55 per cent overlap.

Table 2.52 Descriptive Statistics comparing NS-SEC by SOC2000 and SOC90 codes for BCS model

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Descriptive Statistics comparing NS-SEC by SOC2000 and SOC90 codes | | | | | | | | | |
|  | NS-SEC Social Class of Father when Respondent Child SOC90 | | | | | | | | |
|  | 1.1 | 1.2 | 2 | 3 | 4 | 5 | 6 | 7 | Total |
| N | 85.00 (5.40%) | 132.00 (8.39%) | 315.00 (20.01%) | 178.00 (11.31%) | 157.00 (9.97%) | 261.00 (16.58%) | 199.00 (12.64%) | 247.00 (15.69%) | 1574.00 (100.00%) |
| NS-SEC Social Class of Father when Respondent Child SOC2000 |  |  |  |  |  |  |  |  |  |
| 1.1 | 75 (88.24%) | 1 (0.76%) | 6 (1.90%) | 5 (2.81%) | 2 (1.27%) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) | 89 (5.65%) |
| 1.2 | 2 (2.35%) | 106 (80.30%) | 7 (2.22%) | 3 (1.69%) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) | 118 (7.50%) |
| 2 | 7 (8.24%) | 25 (18.94%) | 251 (79.68%) | 4 (2.25%) | 5 (3.18%) | 1 (0.38%) | 2 (1.01%) | 1 (0.40%) | 296 (18.81%) |
| 3 | 0 (0.00%) | 0 (0.00%) | 8 (2.54%) | 156 (87.64%) | 1 (0.64%) | 7 (2.68%) | 4 (2.01%) | 3 (1.21%) | 179 (11.37%) |
| 4 | 0 (0.00%) | 0 (0.00%) | 29 (9.21%) | 1 (0.56%) | 138 (87.90%) | 1 (0.38%) | 0 (0.00%) | 1 (0.40%) | 170 (10.80%) |
| 5 | 1 (1.18%) | 0 (0.00%) | 5 (1.59%) | 0 (0.00%) | 2 (1.27%) | 243 (93.10%) | 3 (1.51%) | 4 (1.62%) | 258 (16.39%) |
| 6 | 0 (0.00%) | 0 (0.00%) | 8 (2.54%) | 8 (4.49%) | 2 (1.27%) | 0 (0.00%) | 178 (89.45%) | 2 (0.81%) | 198 (12.58%) |
| 7 | 0 (0.00%) | 0 (0.00%) | 1 (0.32%) | 1 (0.56%) | 7 (4.46%) | 9 (3.45%) | 12 (6.03%) | 236 (95.55%) | 266 (16.90%) |

For the RGSC schema in table 2.42, looking at the diagonals all RGSC categories share an overlap of at least 65 per cent. The lowest overlap occurs in RGSC 3NM at 65.79 per cent and the highest overlap occurs in RGSC 3M at 94.22 per cent.

Table 2.53 Descriptive Statistics comparing RGSC by SOC2000 and SOC90 codes for BCS model

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Descriptive Statistics comparing RGSC by SOC2000 and SOC90 codes | | | | | | | |
|  | RGSC Social Class of Father when Respondent Child SOC90 | | | | | | |
|  | 1 | 2 | 3NM | 3M | 4 | 5 | Total |
| N | 123 (7.81%) | 391 (24.84%) | 266 (16.90%) | 536 (34.05%) | 204 (12.96%) | 54 (3.43%) | 1574 (100.00%) |
| RGSC Social Class of Father when Respondent Child SOC2000 |  |  |  |  |  |  |  |
| 1 | 90 (73.17%) | 1 (0.26%) | 1 (0.38%) | 0 (0.00%) | 0 (0.00%) | 0 (0.00%) | 92 (5.84%) |
| 2 | 32 (26.02%) | 352 (90.03%) | 66 (24.81%) | 2 (0.37%) | 6 (2.94%) | 0 (0.00%) | 458 (29.10%) |
| 3NM | 0 (0.00%) | 11 (2.81%) | 175 (65.79%) | 5 (0.93%) | 5 (2.45%) | 0 (0.00%) | 196 (12.45%) |
| 3M | 0 (0.00%) | 21 (5.37%) | 4 (1.50%) | 505 (94.22%) | 45 (22.06%) | 2 (3.70%) | 577 (36.66%) |
| 4 | 1 (0.81%) | 6 (1.53%) | 7 (2.63%) | 19 (3.54%) | 137 (67.16%) | 4 (7.41%) | 174. (11.05%) |
| 5 | 0 (0.00%) | 0 (0.00%) | 13 (4.89%) | 5 (0.93%) | 11 (5.39%) | 48 (88.89%) | 77 (4.89%) |

Table 2.54 Descriptive Statistics comparing CAMSIS by SOC2000 and SOC90 codes by BCS model

|  |  |
| --- | --- |
| CAMSIS2000 | |
| *Mean* | 49.06 |
| *Standard Deviation* | 13.81 |
| CAMSIS90 | |
| *Mean* | 49.38 |
| *Standard Deviation* | 14.52 |
| N | 1574 |

For CAMSIS for SOC 2000 construction, with a base total mean of 49.06 and a standard deviation of 13.81. The SOC 90 construction of CAMSIS has a base mean of 49.38 and a standard deviation of 14.52. CAMSIS measures are remarkably similar across SOC codes.

### Initial Model

Table 2.44 details the deviance, change in deviance, change in degrees of freedom, and McFadden’s Pseudo , AIC, and BIC measures to compare the null model with models of one explanatory variable. Table 2.45 details the null model plus each analytical variable. Table 2.45 displays exact statistics but through a sequential building of the null model with each subsequent independent variable added. The model output uses the reference category of 'Don't Continue Schooling'. The reference category contrasts with continuing schooling. The model is identical in construction to the previous chapter and so details of model production can be found in the prior chapter.

Table 2.55 Model Building Statistics for BCS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d. f. (from Null) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 2114.02 | - | - | - | 2116.02 | 2121.38 |
| Null Model + Educational Attainment | 1985.95 | 128.07 | 2 | 0.06 | 1989.95 | 2000.67 |
| Null Model + Sex | 2090.60 | 23.42 | 2 | 0.01 | 2094.60 | 2105.32 |
| Null Model + Tenure | 2099.44 | 14.58 | 2 | 0.01 | 2103.44 | 2114.17 |
| Null Model + NS-SEC (SOC 2000) | 2059.52 | 54.50 | 8 | 0.04 | 2075.52 | 2118.42 |

Table 2.56 Sequential Model Building Statistics of NS-SEC for BCS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 2114.02 | - | - | - | 2116.02 | 2121.38 |
| Null Model + Educational Attainment | 1985.95 | 128.07 | 2 | 0.06 | 1989.95 | 2000.67 |
| Null Model + Educational Attainment + Sex | 1957.40 | 28.55 | 2 | 0.07 | 1963.40 | 1979.48 |
| Null Model + Educational Attainment + Sex + Tenure | 1951.72 | 5.68 | 2 | 0.08 | 1959.72 | 1981.16 |
| Null Model + Educational Attainment + Sex + Tenure + NS-SEC (SOC 2000) | 1923.13 | 28.59 | 8 | 0.09 | 1945.13 | 2004.11 |

The results of KHB decompositions for each additive model for the BCS cohort are found as total effects in table 2.56 and the summary statistics are provided in table 2.57. The confounding ratio across all models does not increase beyond 1.03. The resulting change in total effects is mostly explained by the mediating additional variables added to the models. Like the NCDS sample, the BCS sample demonstrates that the cofounding ratio and percentage across variables is minor with the exception of housing tenure. Model 3 illustrates that the addition of NS-SEC to the model can attribute 35.61 percent change in total effect size of housing tenure.

Table 2.57 KHB Nested Regression Comparisons for BCS cohort

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Model 1 | | Model 2 | | Model 3 | |
|  |  | **Log Odds** | **SE** | **Log Odds** | **SE** | **Log Odds** | **SE** |
| Educational Attainment | *Reduced* | 1.28\*\*\* | (0.12) | 1.30\*\*\* | (0.12) | 1.30\*\*\* | (0.12) |
| *Full* | 1.30\*\*\* | (0.12) | 1.27\*\*\* | (0.12) | 1.21\*\*\* | (0.12) |
| *Difference* | -0.02 | (0.02) | 0.04\* | (0.02) | 0.09\*\* | (0.04) |
| Sex | *Reduced* |  |  | -0.58\*\*\* | (0.11) | -0.59\*\*\* | (0.11) |
| *Full* |  |  | -0.58\*\*\* | (0.11) | -0.58\*\*\* | (0.11) |
| *Difference* |  |  | -0.00 | (0.01) | -0.01 | (0.03) |
| Housing Tenure | *Reduced* |  |  |  |  | -0.31\*\* | (0.13) |
| *Full* |  |  |  |  | -0.20 | (0.13) |
| *Difference* |  |  |  |  | -0.11\*\* | (0.04) |

Table 2.58 KHB Summary Statistics of Nested Regression Comparisons for BCS cohort

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Model 1 | | | Model 2 | | | Model 3 | | |
|  | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** |
| Educational Attainment | 0.99 | -1.16 | 1.02 | 1.03 | 2.75 | 1.00 | 1.07 | 6.66 | 1.02 |
| Sex |  |  |  | 1.00 | 0.44 | 1.00 | 1.01 | 0.85 | 1.02 |
| Housing Tenure |  |  |  |  |  |  | 1.55 | 35.61 | 1.03 |

The results of the logistic regression model are reported in Table 2.46. Following from prior analysis both log odds, average marginal effects, and quasi-variance statistics are reported.

Table 2.59 Analytical Model for BCS

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | NS-SEC | | | Average Marginal Effects | | Quasi-variance | | |
| Economic Activity: ‘Don’t Continue Schooling’ Reference Category | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **S.E.** | **LB** | **UB** |
| Continue Schooling |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. |  |  |  |  |  |  |  |
| *Five or More O’levels* | 1.21 | (0.12) | \*\*\* | 0.26 | (0.02) |  |  |  |
| Sex |  |  |  |  |  |  |  |  |
| *Female* | Ref. |  |  |  |  |  |  |  |
| *Male* | -0.58 | (0.11) | \*\*\* | -0.13 | (0.02) |  |  |  |
| Housing Tenure |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. |  |  |  |  |  |  |  |
| *Do not Own Home* | -0.20 | (0.13) |  | -0.04 | (0.03) |  |  |  |
| NS-SEC (SOC 2000) |  |  |  |  |  |  |  |  |
| *1.1* | 0.29 | (0.29) |  | 0.06 | (0.05) | 0.25 | -0.27 | 0.86 |
| *1.2* | 0.12 | (0.26) |  | 0.02 | (0.05) | 0.22 | -0.38 | 0.61 |
| *2* | Ref. |  |  |  |  | 0.14 | -0.30 | 0.30 |
| *3* | -0.20 | (0.21) |  | -0.04 | (0.04) | 0.17 | -0.56 | 0.17 |
| *4* | -0.57 | (0.21) | \*\* | -0.12 | (0.05) | 0.16 | -0.93 | -0.21 |
| *5* | -0.70 | (0.19) | \*\*\* | -0.15 | (0.04) | 0.13 | -0.99 | -0.41 |
| *6* | -0.35 | (0.20) |  | -0.07 | (0.04) | 0.15 | -0.69 | -0.01 |
| *7* | -0.50 | (0.19) | \*\* | -0.11 | (0.04) | 0.13 | -0.79 | -0.21 |
| Intercept | 0.63 | (0.15) | \*\*\* |  |  |  |  |  |
| Number of observations | 1574 | | | | | | | |
| McFadden’s | 0.09 | | | | | | | |
| McFadden’s Adjusted Pseudo | 0.08 | | | | | | | |
| Cox-Snell Pseudo | 0.11 | | | | | | | |
| Nagelkerke Pseudo | 0.16 | | | | | | | |
| Tjur’s | 0.13 | | | | | | | |
| AIC | 1945.13 | | | | | | | |
| BIC | 2004.48 | | | | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: BCS [Sweeps 0-5]  Note: Complete Records Analysis | | | | | | | | |

The output from table 2.46 demonstrates that for those individuals who received five or more O’levels compared to those that received less than five O’levels had an increased log odds of continuing schooling. Translated to average marginal effects, this represents an increase of 26 percentage points to continue schooling if individuals received five or more O’levels. This result suggests that prior educational attainment has a substantial impact on a young persons first transition within the BCS. Prior educational attainment has a large influence on a young persons first transition.

Men in the BCS had a decreased log odds of continuing schooling versus women. Translated into average marginal effects, this is a 13 percentage point decrease of continuing schooling for men compared to women. Sex has a sizeable influence on a young persons first transition. There is a sexed divide amongst the BCS cohort that sees men less likely to continue schooling compared to women. This may be due to the structural divisions within the BCS cohort that emphasis men entering employment more readily compared to women.

Housing tenure is not found to be statistically significant and as such will not be substantively interpreted.

Finally moving on to NS-SEC, classes 4, 5 and 7 are statistically significant. Individuals in NS-SEC 4, 5 and 7 compared to the reference category of NS-SEC 2 had a decreased log odds of continuing schooling. Translated into average marginal effects this represents a range of a 11-15 percentage point decrease of continuing schooling. The fact that NS-SEC does not report a monotonic pattern suggests that there are unique phenomena occurring at different points of NS-SEC categories. For NS-SEC 4 and 5, who make up small employers and own account workers as well as lower supervisory and technical occupations there is a qualitative difference between individuals in this category and routine occupations that make up NS-SEC 7. Individuals in NS-SEC 4 and 5 would not necessarily require traditional educational institutions to follow in their parents’ footsteps. Small employers and technical occupations would benefit most from non-traditional occupations and apprenticeship schemes rather than continuing schooling. It would be sensible to attribute the decreased percentage in continuing schooling to these reasons. Members of NS-SEC 7 are from the most disadvantaged NS-SEC category, rationally, entering employment as fast as possible would be a sensible option for individuals that do not have the material support to continue schooling. This would be a sensible attribution of the decreased percentage in continuing schooling for NS-SEC 7 members. Whilst all three NS-SEC categories have a decreased likelihood of continuing schooling, the reasons for doing so may not be identical, which may in part, explain the non-monotonic pattern. If there were one singular reason, then a monotonic decrease (or increase) would be more likely.

To understand these statistics in a more intuitive format, graphs are produced to aid in substantive interpretation. First figure 2.31 provides a coefficient plot of all main analytical variables within the model, in place of statistics produced via table 2.59.

A graph with lines and dots

Description automatically generated

Figure ‎2.30 Coefficient plot of main analytical model

Secondly, figure 2.32 isolates NS-SEC as the only multiple categorical variables by presenting the log odds and quasi-variance statistics. Figure 2.32 demonstrates that quasi-variance statistics provide adjusted smaller bounds compared to the confidence intervals of the log odds presented. As for the NCDS sample, the BCS sample also appears to have smaller adjusted bounds for NS-SEC 4-7 compared to the upper end of the category.

A graph with red and black lines

Description automatically generated

Figure ‎2.31 Log odds versus Quasi-Variance Statistics for BCS model (NS-SEC)

NS-SEC is graphed using both predictive probabilities and average marginal effects in figure 2.33. The predictive probabilities presented in figure 2.32 show that all NS-SEC categories have a rather high predictive probability of continuing schooling though there is a substantive dip starting at NS-SEC 3-5. The average marginal effects report a clear decrease in percentage points for NS-SEC 4, 5, and 7.

A screenshot of a computer screen

Description automatically generated

Figure ‎2.32 Predictive and AMEs of NS-SEC for BCS Model

Figure 2.33 reports the predictive probabilities for educational attainment. There is a clear, pronounced difference in the predicted probabilities of continuing schooling for each educational attainment category. There is close to a 0.3 different between categories.

A white rectangular object with black text

Description automatically generated

Figure ‎2.33 Predictive Margins of Educational Attainment for BCS model

Figure 2.34 reports the predictive probabilities for sex. There is a clear, pronounced difference in the predicted probabilities of continuing schooling for male and female categories of sex.

A screenshot of a computer

Description automatically generated

Figure ‎2.34 Predictive Margins of Sex for BCS model

Figure 2.35 reports the predictive probabilities for housing tenure. The lack of statistical significance in this variable is found in the crossing confidence interval boundaries of the predicted probabilities of each category of housing tenure.

A white rectangular object with black text

Description automatically generated

Figure ‎2.35 Predictive Margins of Housing Tenure for BCS model

### Sensitivity Analysis of Social Stratification Measures using BCS

Following from the sensitivity analysis of social stratification measures conducted in the previous chapter, a duplicate analysis will take place using the BCS cohort to identify any similarities or differences in the substantive interpretation using a different sample. Three social stratification measures: NS-SEC, RGSC, and CAMSIS will be used. The subsequent sensitivity analysis will compare like-for-like models of economic activity, each using a different socio-economic measure. The base model – NS-SEC – is then compared to the CAMSIS and RGSC models. KHB decomposition is employed to assess the level of rescaling across each model for comparative purposes. The following sensitivity analysis will compare models following substantive conclusions. Goodness-of-fit statistics are provided and are assessed via AIC, BIC, and a range of measures.

Three separate multinomial logistic regressions are presented in Table 2.51. The first model has been described at length in the previous section and uses NS-SEC. The second model uses CAMSIS, and the third uses RGSC. These models are all presented using log odds and average marginal effects to enhance interpretation and comparison.

As with the base NS-SEC model above, model building statistics for both RGSC and CAMSIS measures are provided in tables 2.47-2.50.

Table 2.60 Model building statistics of RGSC for BCS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d. f. (from Null) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 2114.02 | - | - | - | 2116.02 | 2121.38 |
| Null Model + Educational Attainment | 1985.95 | 128.07 | 2 | 0.06 | 1989.95 | 2000.67 |
| Null Model + Sex | 2090.60 | 23.42 | 2 | 0.01 | 2094.60 | 2105.32 |
| Null Model + Tenure | 2099.44 | 14.58 | 2 | 0.01 | 2103.44 | 2114.17 |
| Null Model + RGSC (SOC 2000) | 2054.91 | 59.11 | 6 | 0.03 | 2066.91 | 2099.08 |

Table 2.61 Sequential Model Building Statistics of RGSC for BCS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 2114.02 | - | - | - | 2116.02 | 2121.38 |
| Null Model + Educational Attainment | 1985.95 | 128.07 | 2 | 0.06 | 1989.95 | 2000.67 |
| Null Model + Educational Attainment + Sex | 1957.40 | 28.55 | 2 | 0.07 | 1963.40 | 1979.48 |
| Null Model + Educational Attainment + Sex + Tenure | 1951.72 | 5.68 | 2 | 0.08 | 1959.72 | 1981.16 |
| Null Model + Educational Attainment + Sex + Tenure + RGSC (SOC 2000) | 1920.92 | 30.80 | 6 | 0.09 | 1938.92 | 1987.17 |

Table 2.62 Model building statistics of CAMSIS for BCS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d. f. (from Null) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 2114.02 | - | - | - | 2116.02 | 2121.38 |
| Null Model + Educational Attainment | 1985.95 | 128.07 | 2 | 0.06 | 1989.95 | 2000.67 |
| Null Model + Sex | 2090.60 | 23.42 | 2 | 0.01 | 2094.60 | 2105.32 |
| Null Model + Tenure | 2099.44 | 14.58 | 2 | 0.01 | 2103.44 | 2114.17 |
| Null Model + CAMSIS (SOC 2000) | 2054.16 | 59.86 | 2 | 0.03 | 2058.16 | 2068.89 |

Table 2.63 Sequential Model Building Statistics of CAMSIS for BCS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 2114.02 | - | - | - | 2116.02 | 2121.38 |
| Null Model + Educational Attainment | 1985.95 | 128.07 | 2 | 0.06 | 1989.95 | 2000.67 |
| Null Model + Educational Attainment + Sex | 1957.40 | 28.55 | 2 | 0.07 | 1963.40 | 1979.48 |
| Null Model + Educational Attainment + Sex + Tenure | 1951.72 | 5.68 | 2 | 0.08 | 1959.72 | 1981.16 |
| Null Model + Educational Attainment + Sex + Tenure + CAMSIS (SOC 2000) | 1926.65 | 25.07 | 2 | 0.09 | 1936.65 | 1963.46 |

Tables 2.63 and 2.64 detail the KHB method comparing each social stratification measure-based model using total effect sizes in table 2.63 and summary statistics in table 2.64. All three models present near identical findings – rescaling is a minor attribute tot eh changing effect size within subsequent models. Housing tenure remains mediated by the social stratification measures used across models 1-3. The largest mediated effect comes from CAMSIS in model 3. The total effect size of housing tenure increased by 48.96 percent due to the inclusion of CAMSIS within the model. This echoes the findings of the NCDS sample – whereby CAMSIS was also the largest mediator across social stratification models for housing tenure also. Though in the BCS sample, the increase is even larger.

Table 2.64 method of Nested Regression Models for Social Stratification Analysis for BCS cohort

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | Model 1 | | Model 2 | | Model 3 | |
|  |  | **Log Odds** | **SE** | **Log Odds** | **SE** | **Log Odds** | **SE** |
| Educational Attainment | *Reduced* | 1.30\*\*\* | (0.12) | 1.29\*\*\* | (0.12) | 1.29\*\*\* | (0.12) |
| *Full* | 1.21\*\*\* | (0.12) | 1.19\*\*\* | (0.12) | 1.16\*\*\* | (0.12) |
| *Difference* | 0.09\*\* | (0.04) | 0.10\*\* | (0.04) | 0.12\*\*\* | (0.04) |
| Sex | *Reduced* | -0.59\*\*\* | (0.11) | -0.59\*\*\* | (0.11) | -0.59\*\*\* | (0.11) |
| *Full* | -0.58\*\*\* | (0.11) | -0.58\*\*\* | (0.11) | -0.58\*\*\* | (0.11) |
| *Difference* | -0.01 | (0.03) | -0.01 | (0.03) | -0.01 | (0.03) |
| Housing Tenure | *Reduced* | -0.31\*\* | (0.13) | -0.31\*\* | (0.13) | -0.30\*\* | (0.13) |
| *Full* | -0.20 | (0.13) | -0.19 | (0.13) | -0.16 | (0.13) |
| *Difference* | -0.11\*\* | (0.04) | -0.12\*\*\* | (0.04) | -0.15\*\*\* | (0.04) |

Table 2.65 KHB method Summary Statistics for Nested Regression Models for Social Stratification Analysis for BCS cohort

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Model 1 | | | Model 2 | | | Model 3 | | |
|  | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** |
| Educational Attainment | 1.07 | 6.66 | 1.02 | 1.09 | 8.03 | 1.02 | 1.11 | 9.65 | 1.02 |
| Sex | 1.01 | 0.85 | 1.02 | 1.01 | 1.45 | 1.02 | 1.01 | 0.90 | 1.01 |
| Housing Tenure | 1.55 | 35.61 | 1.03 | 1.62 | 38.35 | 1.02 | 1.96 | 48.96 | 1.01 |

Moving on to interpretation of the sensitivity analysis of social stratification in table 2.51, some differences are identified. Model one will not be substantively interpreted independently due to it being discussed previously. It will instead be used as a reference model to compare to. All non-social stratification-based measures in the analytical models provide near-identical substantive findings: educational attainment, sex, and housing tenure.

A comparison of the NS-SEC and RGSC models identifies a remarkable substantive similarity, something that is shared between the NCDS and BCS models. Just as NS-SEC has a decreased probability for NS-SEC 4, 5 and 7 compared to its reference category of NS-SEC 2, the RGSC model also sees RGSC 3M-5 have a decreased log odds compared to RGSC 2 for continuing schooling. The RGSC model explicitly demonstrates a manual/non-manual divide – again, like the NCDS model. Whilst there is no substantive difference within manual occupation social origins in RGSC, there is a substantive difference between manual and non-manual occupations with up to a 15 per cent decreased probability of continuing schooling compared to the RGSC 2 peers.

Moving on to the CAMSIS model demonstrates just as with the NCDS analysis that whilst statistically significant CAMSIS holds no substantive significance – unlike the NS-SEC and RGSC models.

All three models have near identical . Overall, the models suggest a small amount of the overall variance is explained by each model. AIC and BIC statistics demonstrate a very marginal difference between all three models. Unsurprisingly the CAMSIS model is most favoured by BIC and AIC statistics, though the difference is tiny. As with the NCDS analysis, going forward the NS-SEC model is selected for further investigation.

Table 2.66 Sensitivity analysis of social stratification measures for BCS model

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | NS-SEC | | | Average Marginal Effects | | Quasi-variance | | | RGSC | | | Average Marginal Effects | | Quasi-variance | | | CAMSIS | | | Average Marginal Effects | |
| Economic Activity: ‘Don’t Continue Schooling’ Reference Category | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **S.E.** | **LB** | **UB** | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **S.E.** | **LB** | **UB** | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** |
| Continue Schooling |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Five or More O’levels* | 1.21 | (0.12) | \*\*\* | 0.26 | (0.02) |  |  |  | 1.19 | (0.12) | \*\*\* | 0.26 | (0.02) |  |  |  | 1.16 | (0.12) | \*\*\* | 0.25 | (0.02) |
| Sex |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. |  |  |  |  |  |  |  | (.) | (.) |  | (.) | (.) |  |  |  | (.) | (.) |  | (.) | (.) |
| *Male* | -0.58 | (0.11) | \*\*\* | -0.13 | (0.02) |  |  |  | -0.58 | (0.11) | \*\*\* | -0.12 | (0.02) |  |  |  | -0.58 | (0.11) | \*\*\* | -0.13 | (0.02) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. |  |  |  |  |  |  |  | (.) | (.) |  | (.) | (.) |  |  |  | (.) | (.) |  | (.) | (.) |
| *Do not Own Home* | -0.20 | (0.13) |  | -0.04 | (0.03) |  |  |  | -0.19 | (0.13) |  | -0.04 | (0.03) |  |  |  | -0.15 | (0.13) |  | -0.03 | (0.03) |
| NS-SEC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1.1* | 0.29 | (0.29) |  | 0.06 | (0.05) | 0.25 | -0.27 | 0.86 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1.2* | 0.12 | (0.26) |  | 0.02 | (0.05) | 0.22 | -0.38 | 0.61 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *2* | Ref. |  |  |  |  | 0.14 | -0.30 | 0.30 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *3* | -0.20 | (0.21) |  | -0.04 | (0.04) | 0.17 | -0.56 | 0.17 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *4* | -0.57 | (0.21) | \*\* | -0.12 | (0.05) | 0.16 | -0.93 | -0.21 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *5* | -0.70 | (0.19) | \*\*\* | -0.15 | (0.04) | 0.13 | -0.99 | -0.41 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *6* | -0.35 | (0.20) |  | -0.07 | (0.04) | 0.15 | -0.69 | -0.01 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *7* | -0.50 | (0.19) | \*\* | -0.11 | (0.04) | 0.13 | -0.79 | -0.21 |  |  |  |  |  |  |  |  |  |  |  |  |  |
| RGSC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1* |  |  |  |  |  |  |  |  | -0.01 | (0.27) |  | -0.00 | (0.05) | 0.25 | -0.58 | 0.56 |  |  |  |  |  |
| *2* |  |  |  |  |  |  |  |  | Ref. |  |  |  |  | 0.11 | -0.25 | 0.25 |  |  |  |  |  |
| *3NM* |  |  |  |  |  |  |  |  | -0.24 | (0.19) |  | -0.05 | (0.04) | 0.16 | -0.60 | 0.12 |  |  |  |  |  |
| *3M* |  |  |  |  |  |  |  |  | -0.71 | (0.14) | \*\*\* | -0.15 | (0.03) | 0.09 | -0.91 | -0.51 |  |  |  |  |  |
| *4* |  |  |  |  |  |  |  |  | -0.54 | (0.20) | \*\* | -0.11 | (0.04) | 0.16 | -0.91 | -0.17 |  |  |  |  |  |
| *5* |  |  |  |  |  |  |  |  | -0.61 | (0.26) | \* | -0.13 | (0.06) | 0.24 | -1.16 | -0.06 |  |  |  |  |  |
| CAMSIS (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | 0.02 | (0.00) | \*\*\* | 0.00 | (0.00) |
| Intercept | 0.63 | (0.15) | \*\*\* |  |  |  |  |  | 0.71 | (0.13) | \*\*\* |  |  |  |  |  | -0.71 | (0.22) | \*\*\* |  |  |
| Number of observations | 1574 | | | | | | | | 1574 | | | | | | | | 1574 | | | | |
| McFadden’s | 0.09 | | | | | | | | 0.09 | | | | | | | | 0.09 | | | | |
| McFadden’s Adjusted Pseudo | 0.08 | | | | | | | | 0.08 | | | | | | | | 0.08 | | | | |
| Cox-Snell Pseudo | 0.11 | | | | | | | | 0.11 | | | | | | | | 0.11 | | | | |
| Nagelkerke Pseudo | 0.16 | | | | | | | | 0.16 | | | | | | | | 0.16 | | | | |
| Tjur’s | 0.13 | | | | | | | | 0.13 | | | | | | | | 0.13 | | | | |
| AIC | 1945.13 | | | | | | | | 1938.92 | | | | | | | | 1936.65 | | | | |
| BIC | 2004.48 | | | | | | | | 1987.17 | | | | | | | | 1963.46 | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: BCS [Sweeps 0-5]  Note: Sensitivity Analysis of Social Stratification Measures | | | | | | | | | | | | | | | | | | | | | |

A graph with lines and dots

Description automatically generated with medium confidence

Figure ‎2.36 Combined coefficient plot of SOC 2000 social stratification measures

The coefficients of all three social stratification models are plotted in figure 2.37. Each model has near identical coefficients and confidence intervals for each analytical variable shared across each model. The overall general trend documented by NS-SEC and RGSC in earlier interrpetation repeats itself in a visual format. CAMSIS remains near zero.

The log odds and quasi-variance statistics are graphed visually for RGSC in figure 2.38. From this graph a manual/non-manual divide is evident whereby individuals from manual social origins have a decreased log odds of continuing schooling compared to the reference category of RGSC 2. Figure 2.38 also directly compares the log odds and quasi-variance statistics of both NS-SEC and RGSC models. There is a shared general pattern that demonstrates compared to each respective reference category, the lower end of each schema has a decreased log odds of continuing schooling.

A diagram of a graph

Description automatically generated with medium confidence

Figure ‎2.37 Comparison of log odds versus quasi-variance statistics of NS-SEC and RGSC measures for BCS model

Figure 2.39 depicts the predicted probability at means of economic activity alongside the average marginal effects of each social stratification measure. All graphs are represented using the same common y axis to aid interpretation. Starting with RGSC, with respect to predicted probabilities there is a decrease from 3NM to 3M which represents a key distinction between manual and non-manual occupations. The average marginal effects graph also demonstrates that there is little distinction within manual occupation categories – so long as an individual is a member of a manual occupation, they have a near flat penalty of continuing schooling compared to the reference category of RGSC 2.

Whilst the predictive probability of CAMSIS demonstrates a linear positive monotonic increase the average marginal effects demonstrate a complete flatline of 0 per cent.

Once more there is evidence that the NS-SEC and RGSC models produce a similar substantive interpretation compared to CAMSIS which provides a distinct substantive pattern. However, it would be false to conclude that the NS-SEC and RGSC models offer identical substantive interpretation. The latter of these models demonstrates a clear and distinct manual/non-manual pattern that is obfuscated by NS-SECs construction.

A screenshot of a computer screen

Description automatically generated

Figure ‎2.38 Comparison of Predictive and AMEs for each social stratification measure for BCS mode

#### Discussion and Conclusion

The overall substantive findings of the BCS sensitivity analysis of social stratification measures echoes the findings of the NCDS sensitivity analysis of the same measures. Two key stories develop. The first relates to the similarity of the NS-SEC and RGSC models – both provide similar substantive interpretations whether looking at log odds, average marginal effects, quasi-variance statistics, or predictive probabilities at mean of economic activity. The one key difference between the two social class measures is that by its nature RGSC promotes a story of manual/non-manual division which NS-SEC does not due to its design and explicitly rejection of such a dichotomy in British society.

The second story relates to the dissimilarity between CAMSIS and the other two social stratification measures. Once more this is a finding repeated from the NCDS sensitivity analysis and once more confirms that whilst all social stratification measures, CAMSIS appears to measure something distinct from social class measures of stratification.

This sensitivity analysis was an exploration of social stratification measures and a duplication-based analysis of the NCDS sensitivity analysis using the same stratification measures. The findings within the BCS cohort confirm and strengthen the points made in the discussion and conclusions of the NCDS cohort.

### SOC Code Sensitivity analysis using BCS

Following from the sensitivity analysis of social stratification measures, a sensitivity analysis of SOC codes will now proceed. Another sensitivity analysis will be conducted comparing the measure of NS-SEC under two different constructions. The first will be NS-SEC constructed using SOC 2000 codes – the base model used previously. The second, will use NS-SEC constructed using SOC 90 codes. These two models will be compared to assess any similarities and differences regarding their substantive effects. Goodness-of-fit statistics will also be assessed to determine the best fit model. A comparison of SOC 2000 and SOC 90 codes for both RGSC and CAMSIS models follows from this initial NS-SEC model to estimate any differences in substantive interpretation if using a different social stratification measure.

The following tables 2.52-2.53 detail model building statistics for NS-SEC under a SOC 90 construction.

Table 2.67 Model building statistics of NS-SEC SOC 90 for BCS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d. f. (from Null) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 2114.02 | - | - | - | 2116.02 | 2121.38 |
| Null Model + Educational Attainment | 1985.95 | 128.07 | 2 | 0.06 | 1989.95 | 2000.67 |
| Null Model + Sex | 2090.60 | 23.42 | 2 | 0.01 | 2094.60 | 2105.32 |
| Null Model + Tenure | 2099.44 | 14.58 | 2 | 0.01 | 2103.44 | 2114.17 |
| Null Model + NS-SEC (SOC 90) | 2057.04 | 56.98 | 8 | 0.03 | 2073.04 | 2115.93 |

Table 2.68 Sequential Model Statistics of NS-SEC SOC 90 for BCS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 2114.02 | - | - | - | 2116.02 | 2121.38 |
| Null Model + Educational Attainment | 1985.95 | 128.07 | 2 | 0.06 | 1989.95 | 2000.67 |
| Null Model + Educational Attainment + Sex | 1957.40 | 28.55 | 2 | 0.07 | 1963.40 | 1979.48 |
| Null Model + Educational Attainment + Sex + Tenure | 1951.72 | 5.68 | 2 | 0.08 | 1959.72 | 1981.16 |
| Null Model + Educational Attainment + Sex + Tenure + NS-SEC (SOC 90) | 1920.97 | 30.75 | 8 | 0.09 | 1942.97 | 2001.94 |

A KHB decomposition analysis is presented in tables 2.68 and 2.69 to provide the change in total effect sizes and summary statistics for each model. Model 1 uses a SOC 2000 construction of NS-SEC and model 2 uses a SOC 90 construction of NS-SEC. Similarly to the NCDS sample when comparing NS-SEC SOC construction measures, both models report near identical results, though the confounding percentage for housing tenure is lower in model 2 than it is in model 1. That is to say the mediating effect of NS-SEC with a SOC 90 construction is less than it is for NS-SEC with a SOC 2000 construction.

Table 2.69 A Comparison of SOC NS-SEC measures using the KHB method for BCS cohort

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Model 1 | | Model 2 | |
|  |  | **Log Odds** | **SE** | **Log Odds** | **SE** |
| Educational Attainment | *Reduced* | 1.30\*\*\* | (0.12) | 1.29\*\*\* | (0.12) |
| *Full* | 1.21\*\*\* | (0.12) | 1.21\*\*\* | (0.12) |
| *Difference* | 0.09\*\* | (0.04) | 0.09\* | (0.04) |
| Sex | *Reduced* | -0.59\*\*\* | (0.11) | -0.59\*\*\* | (0.11) |
| *Full* | -0.58\*\*\* | (0.11) | -0.58\*\*\* | (0.11) |
| *Difference* | -0.01 | (0.03) | -0.01 | (0.03) |
| Housing Tenure | *Reduced* | -0.31\*\* | (0.13) | -0.31\* | (0.13) |
| *Full* | -0.20 | (0.13) | -0.22 | (0.13) |
| *Difference* | -0.11\*\* | (0.04) | -0.09\*\* | (0.04) |

Table 2.70 KHB Summary statistics Comparing SOC NS-SEC models for BCS cohort

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Model 1 | | | Model 2 | | |
|  | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** |
| Educational Attainment | 1.07 | 6.66 | 1.02 | 1.07 | 6.61 | 1.02 |
| Sex | 1.01 | 0.85 | 1.02 | 1.02 | 1.97 | 1.02 |
| Housing Tenure | 1.55 | 35.61 | 1.03 | 1.44 | 30.38 | 1.04 |

#### Measuring SOC Codes

Table 2.54 displays a comparison of the NS-SEC SOC 2000 construction model and the SOC 90 construction model. Log odds, average marginal effects, and quasi-variance statistics are provided for ease of interpretation. Unsurprisingly educational attainment, housing tenure, and sex do not differ between the two models. Both SOC constructions find NS-SEC 4, 5 and 7 statistically significant. Though the SOC 90 model also finds NS-SEC 6 statistically significant. Whilst both SOC constructions provide a similar substantive interpretation of NS-SEC, the SOC 90 construction has a slightly different range of average marginal effects in comparison to the SOC 2000 construction. Where the average marginal effects for SOC 2000 lie between 11-15 per cent, the SOC 90 construction has average marginal effects that lie between 9-16 per cent. Whilst not changing the existing substantive pattern, small differences do in fact exist between the two constructions of NS-SEC.

Table 2.71 Comparison of SOC measures for NS-SEC for BCS model

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | NS-SEC (SOC 2000) | | | Average Marginal Effects | | Quasi-variance | | | NS-SEC (SOC 90) | | | Average Marginal Effects | | Quasi-variance | | |
| Economic Activity: ‘Don’t Continue Schooling’ Reference Category | Coef. | S.E. | Sig. | Prob. | S.E. | S.E. | LB | UB | Coef. | S.E. | Sig. | Prob. | S.E. | S.E. | LB | UB |
| Continue Schooling |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Less than five O’levels | Ref. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Five or More O’levels | 1.21 | (0.12) | \*\*\* | 0.26 | (0.02) |  |  |  | 1.21 | (0.12) | \*\*\* | 0.26 | (0.02) |  |  |  |
| Sex |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Female | Ref. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Male | -0.58 | (0.11) | \*\*\* | -0.13 | (0.02) |  |  |  | -0.58 | (0.11) | \*\*\* | -0.12 | (0.02) |  |  |  |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Own Home | Ref. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Do not Own Home | -0.20 | (0.13) |  | -0.04 | (0.03) |  |  |  | -0.22 | (0.13) |  | -0.05 | (0.03) |  |  |  |
| NS-SEC |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1.1 | 0.29 | (0.29) |  | 0.06 | (0.05) | 0.25 | -0.27 | 0.86 | 0.37 | (0.30) |  | 0.07 | (0.05) | 0.27 | -0.22 | 0.97 |
| 1.2 | 0.12 | (0.26) |  | 0.02 | (0.05) | 0.22 | -0.38 | 0.61 | -0.13 | (0.24) |  | -0.03 | (0.05) | 0.20 | -0.58 | 0.32 |
| 2 | Ref. |  |  |  |  | 0.14 | -0.30 | 0.30 |  |  |  |  |  | 0.13 | -0.29 | 0.29 |
| 3 | -0.20 | (0.21) |  | -0.04 | (0.04) | 0.17 | -0.56 | 0.17 | -0.25 | (0.21) |  | -0.05 | (0.04) | 0.17 | -0.62 | 0.12 |
| 4 | -0.57 | (0.21) | \*\* | -0.12 | (0.05) | 0.16 | -0.93 | -0.21 | -0.69 | (0.21) | \*\* | -0.15 | (0.04) | 0.17 | -1.06 | -0.32 |
| 5 | -0.70 | (0.19) | \*\*\* | -0.15 | (0.04) | 0.13 | -0.99 | -0.41 | -0.74 | (0.18) | \*\*\* | -0.16 | (0.04) | 0.13 | -1.04 | -0.45 |
| 6 | -0.35 | (0.20) |  | -0.07 | (0.04) | 0.15 | -0.69 | -0.01 | -0.42 | (0.20) | \* | -0.09 | (0.04) | 0.15 | -0.76 | -0.09 |
| 7 | -0.50 | (0.19) | \*\* | -0.11 | (0.04) | 0.13 | -0.79 | -0.21 | -0.47 | (0.19) | \* | -0.10 | (0.04) | 0.14 | -0.77 | -0.17 |
| Intercept | 0.63 | (0.15) | \*\*\* |  |  |  |  |  | 0.73 | (0.15) | \*\*\* |  |  |  |  |  |
| Number of observations | 1574 | | | | | | | | 1574 | | | | | | | |
| McFadden’s | 0.09 | | | | | | | | 0.09 | | | | | | | |
| McFadden’s Adjusted Pseudo | 0.08 | | | | | | | | 0.08 | | | | | | | |
| Cox-Snell Pseudo | 0.11 | | | | | | | | 0.11 | | | | | | | |
| Nagelkerke Pseudo | 0.16 | | | | | | | | 0.16 | | | | | | | |
| Tjur’s | 0.13 | | | | | | | | 0.13 | | | | | | | |
| AIC | 1945.13 | | | | | | | | 1942.97 | | | | | | | |
| BIC | 2004.48 | | | | | | | | 2001.94 | | | | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: BCS [Sweeps 0-5]  Note: Sensitivity Analysis of SOC | | | | | | | | | | | | | | | | |

Both SOC 2000 and SOC 90 measures of NS-SEC share similar log odds and standard errors displayed in figure 2.41 which provides a direct compariosn in log odds and quasi-variance statsitics for both SOC constructions of NS-SEC.

A graph with red lines and black text

Description automatically generated

Figure ‎2.40 Comparison of log odds versus quasi-variance statistics for NS-SEC SOC Codes (BCS model)

Figure 2.42 details the predicted probability at mean of economic activity and the average marginal effects with NS-SEC 2 as the reference category for both NS-SEC constructions. This figure once again displays the relative similarity between both measures – minor differences exist but the substantive interpretation remains the same across both SOC measures.

A screenshot of a computer screen

Description automatically generated

Figure ‎2.41 Comparison of Predictive and AMEs for NS-SEC SOC Codes for BCS Model

Moving on to the comparison of other social stratification measures, starting with RGSC. Table 2.55 details the model statistics for RGSC SOC 90. The SOC 90 construction of RGSC has a reduction of 193.10 deviance from the null model. The measures across SOC 2000 and SOC 90 RGSC are almost identical.

Table 2.72 Model building statistics of RGSC SOC 90 for BCS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d. f. (from Null) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 2114.02 | - | - | - | 2116.02 | 2121.38 |
| Null Model + Educational Attainment | 1985.95 | 128.07 | 2 | 0.06 | 1989.95 | 2000.67 |
| Null Model + Sex | 2090.60 | 23.42 | 2 | 0.01 | 2094.60 | 2105.32 |
| Null Model + Tenure | 2099.44 | 14.58 | 2 | 0.01 | 2103.44 | 2114.17 |
| Null Model + RGSC (SOC 90) | 2057.67 | 56.35 | 6 | 0.03 | 2069.67 | 2101.84 |

Table 2.73 Sequential Model Statistics of RGSC SOC 90 for BCS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 2114.02 | - | - | - | 2116.02 | 2121.38 |
| Null Model + Educational Attainment | 1985.95 | 128.07 | 2 | 0.06 | 1989.95 | 2000.67 |
| Null Model + Educational Attainment + Sex | 1957.40 | 28.55 | 2 | 0.07 | 1963.40 | 1979.48 |
| Null Model + Educational Attainment + Sex + Tenure | 1951.72 | 5.68 | 2 | 0.08 | 1959.72 | 1981.16 |
| Null Model + Educational Attainment + Sex + Tenure + RGSC (SOC 90) | 1923.03 | 28.69 | 6 | 0.09 | 1941.03 | 1989.28 |

A KHB decomposition analysis is presented in tables 2.73 and 2.74 to provide the change in total effect sizes and summary statistics for each model. Model 1 uses a SOC 2000 construction of RGSC and model 2 uses a SOC 90 construction of RGSC. Similarly to the NCDS sample when comparing NS-SEC SOC construction measures, both models report near identical results. One difference is that unlike prior SOC comparisons using the BCS sample (NS-SEC) and other samples using a similar analytical model (NCDS) the confounding percentage for housing tenure is not substantively less for SOC 90 constructions of RGSC compared to SOC 2000 constructions of RGSC.

Table 2.74 A Comparison of SOC RGSC measures using the KHB method for BCS cohort

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Model 1 | | Model 2 | |
|  |  | **Log Odds** | **SE** | **Log Odds** | **SE** |
| Educational Attainment | *Reduced* | 1.29\*\*\* | (0.12) | 1.29\*\*\* | (0.12) |
| *Full* | 1.19\*\*\* | (0.12) | 1.20\*\*\* | (0.12) |
| *Difference* | 0.10\*\* | (0.04) | 0.09\*\*\* | (0.03) |
| Sex | *Reduced* | -0.59\*\*\* | (0.11) | -0.59\*\*\* | (0.11) |
| *Full* | -0.58\*\*\* | (0.11) | -0.58\*\*\* | (0.11) |
| *Difference* | -0.01 | (0.03) | -0.02 | (0.03) |
| Housing Tenure | *Reduced* | -0.31\*\* | (0.13) | -0.31\*\* | (0.13) |
| *Full* | -0.19 | (0.13) | -0.19 | (0.13) |
| *Difference* | -0.12\*\*\* | (0.04) | -0.12\*\*\* | (0.04) |

Table 2.75 KHB Summary statistics Comparing SOC RGSC models for BCS cohort

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Model 1 | | | Model 2 | | |
|  | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** |
| Educational Attainment | 1.09 | 8.03 | 1.02 | 1.08 | 7.30 | 1.02 |
| Sex | 1.01 | 1.45 | 1.02 | 1.03 | 2.55 | 1.02 |
| Housing Tenure | 1.62 | 38.35 | 1.02 | 1.65 | 39.27 | 1.02 |

Like NS-SEC comparisons, the RGSC models depicted in table 2.57 provide identical substantive interpretation for all other analytical variables. Both SOC constructions of RGSC find RGSC 3M-5 to be statistically significant. For RGSC 3M and 4 the substantive interpretation in terms of average marginal effects does not differ by more than 1 per cent. RGSC 5 on the other hand does have a substantively large difference between SOC 2000 and SOC 90 models. Whilst both demonstrates that individuals in social origins RGSC 5 have a decreased log odds of continuing schooling compared to the reference category of RGSC 2 the SOC 2000 model translates this to an average marginal effect of 13 per cent decreased probability whereas the SOC 90 model translates this to an 18 per cent decreased probability. Whilst the same overall substantive picture emerges the 5 per cent decrease between SOC construction of the same measure presents a real substantive difference between the use of the two measures.

Table 2.76 Comparison of RGSC SOC for BCS Model

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | RGSC (SOC 2000) | | | Average Marginal Effects | | Quasi-variance | | | RGSC (SOC 90) | | | Average Marginal Effects | | Quasi-variance | | |
| Economic Activity: ‘Don’t Continue Schooling’ Reference Category | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **S.E.** | **LB** | **UB** | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **S.E.** | **LB** | **UB** |
| Continue Schooling |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Five or More O’levels* | 1.19 | (0.12) | \*\*\* | 0.26 | (0.02) |  |  |  | 1.19 | (0.12) | \*\*\* | 0.26 | (0.02) |  |  |  |
| Sex |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Male* | -0.58 | (0.11) | \*\*\* | -0.12 | (0.02) |  |  |  | -0.57 | (0.11) | \*\*\* | -0.12 | (0.02) |  |  |  |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *Do not Own Home* | -0.19 | (0.13) |  | -0.04 | (0.03) |  |  |  | -0.19 | (0.13) |  | -0.04 | (0.03) |  |  |  |
| RGSC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| *1* | -0.01 | (0.27) |  | -0.00 | (0.05) | 0.25 | -0.58 | 0.56 | -0.10 | (0.24) |  | -0.02 | (0.05) | 0.21 | -0.59 | 0.39 |
| *2* | Ref. |  |  |  |  | 0.11 | -0.25 | 0.25 | Ref. |  |  |  |  | 0.12 | -0.27 | 0.27 |
| *3NM* | -0.24 | (0.19) |  | -0.05 | (0.04) | 0.16 | -0.60 | 0.12 | -0.19 | (0.18) |  | -0.04 | (0.04) | 0.14 | -0.50 | 0.12 |
| *3M* | -0.71 | (0.14) | \*\*\* | -0.15 | (0.03) | 0.09 | -0.91 | -0.51 | -0.68 | (0.15) | \*\*\* | -0.15 | (0.03) | 0.09 | -0.89 | -0.47 |
| *4* | -0.54 | (0.20) | \*\* | -0.11 | (0.04) | 0.16 | -0.91 | -0.17 | -0.57 | (0.19) | \*\* | -0.12 | (0.04) | 0.15 | -0.91 | -0.23 |
| *5* | -0.61 | (0.26) | \* | -0.13 | (0.06) | 0.24 | -1.16 | -0.06 | -0.85 | (0.31) | \*\* | -0.18 | (0.07) | 0.29 | -1.51 | -0.19 |
| Intercept | 0.71 | (0.13) | \*\*\* |  |  |  |  |  | 0.70 | (0.14) | \*\*\* |  |  |  |  |  |
| Number of observations | 1574 | | | | | | | | 1574 | | | | | | | |
| McFadden’s | 0.09 | | | | | | | | 0.09 | | | | | | | |
| McFadden’s Adjusted Pseudo | 0.08 | | | | | | | | 0.08 | | | | | | | |
| Cox-Snell Pseudo | 0.11 | | | | | | | | 0.11 | | | | | | | |
| Nagelkerke Pseudo | 0.16 | | | | | | | | 0.16 | | | | | | | |
| Tjur’s | 0.13 | | | | | | | | 0.13 | | | | | | | |
| AIC | 1938.92 | | | | | | | | 1941.03 | | | | | | | |
| BIC | 1987.17 | | | | | | | | 1989.28 | | | | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: BCS [Sweeps 0-5]  Note: Sensitivity Analysis of SOC RGSC | | | | | | | | | | | | | | | | |

Looking at figure 2.44, log odds and quasi-variance statistics for both SOC constructions of RGSC measures. Both measures provide a clear overall substantive picture of a manual/non-manual divide within the BCS cohort in terms of continuing schooling or not. Figure 2.45 goes further and details the predictive probabilities at means of economic activity and the average marginal effects for RGSC SOC measures – the same substantive picture is drawn from this figure also.

A diagram of a graph

Description automatically generated with medium confidence

Figure ‎2.43 Comparison of log odds versus quasi-variance statistics for RGSC SOC Codes (BCS model)

A screenshot of a computer screen

Description automatically generated

Figure ‎2.44 Comparison of Predictive and AMEs for RGSC SOC Codes for BCS Model

Finally, now moving on to the CAMSIS SOC constructions of social stratification. Table 2.58 and table 2.59 detail the model statistics of CAMSIS for SOC 90. CAMSIS measures share a near identical measure statistics.

Table 2.77 Model building statistics of CAMSIS SOC 90 for BCS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Null) | d. f. (from Null) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 2114.02 | - | - | - | 2116.02 | 2121.38 |
| Null Model + Educational Attainment | 1985.95 | 128.07 | 2 | 0.06 | 1989.95 | 2000.67 |
| Null Model + Sex | 2090.60 | 23.42 | 2 | 0.01 | 2094.60 | 2105.32 |
| Null Model + Tenure | 2099.44 | 14.58 | 2 | 0.01 | 2103.44 | 2114.17 |
| Null Model + CAMSIS (SOC 90) | 2058.23 | 55.79 | 2 | 0.03 | 2062.23 | 2072.95 |

Table 2.78 Sequential Model Statistics of CAMSIS SOC 90 for BCS model

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Outcome Variable: Economic Activity | Deviance | Deviance (from Previous) | d. f. (from Previous) | McFadden’s Pseudo | AIC | BIC |
| Null Model | 2114.02 | - | - | - | 2116.02 | 2121.38 |
| Null Model + Educational Attainment | 1985.95 | 128.07 | 2 | 0.06 | 1989.95 | 2000.67 |
| Null Model + Educational Attainment + Sex | 1957.40 | 28.55 | 2 | 0.07 | 1963.40 | 1979.48 |
| Null Model + Educational Attainment + Sex + Tenure | 1951.72 | 5.68 | 2 | 0.08 | 1959.72 | 1981.16 |
| Null Model + Educational Attainment + Sex + Tenure + CAMSIS (SOC 90) | 1928.23 | 23.49 | 2 | 0.09 | 1938.23 | 1965.04 |

A KHB decomposition analysis is presented in tables 2.78 and 2.79 to provide the change in total effect sizes and summary statistics for each model. Model 1 uses a SOC 2000 construction of CAMSIS and model 2 uses a SOC 90 construction of CAMSIS. Similarly to the NCDS sample when comparing NS-SEC SOC construction measures, both models report near identical results. One difference is that unlike prior SOC comparisons using the BCS sample (NS-SEC) and other samples using a similar analytical model (NCDS) the confounding percentage for housing tenure is not substantively less for SOC 90 constructions of CAMSIS compared to SOC 2000 constructions of CAMSIS. In this regard the CAMSIS constructions using different SOC codes echo results from the RGSC KHB decomposition analysis for the BCS cohort.

Table 2.79 A Comparison of SOC CAMSIS measures using the KHB method for BCS cohort

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  | Model 1 | | Model 2 | |
|  |  | **Log Odds** | **SE** | **Log Odds** | **SE** |
| Educational Attainment | *Reduced* | 1.29\*\*\* | (0.12) | 1.29\*\*\* | (0.12) |
| *Full* | 1.16\*\*\* | (0.12) | 1.17\*\*\* | (0.12) |
| *Difference* | 0.12\*\*\* | (0.04) | 0.12\*\*\* | (0.04) |
| Sex | *Reduced* | -0.59\*\*\* | (0.11) | -0.59\*\*\* | (0.11) |
| *Full* | -0.58\*\*\* | (0.11) | -0.59\*\*\* | (0.11) |
| *Difference* | -0.01 | (0.03) | -0.00 | (0.03) |
| Housing Tenure | *Reduced* | -0.30\*\* | (0.13) | -0.30\* | (0.13) |
| *Full* | -0.16 | (0.13) | -0.16 | (0.13) |
| *Difference* | -0.15\*\*\* | (0.04) | -0.14\*\*\* | (0.04) |

Table 2.80 KHB Summary statistics Comparing SOC CAMSIS models for BCS cohort

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Model 1 | | | Model 2 | | |
|  | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** | **Confounding ratio** | **Confounding percentage** | **Rescaling factor** |
| Educational Attainment | 1.11 | 9.65 | 1.02 | 1.10 | 9.03 | 1.02 |
| Sex | 1.01 | 0.90 | 1.01 | 1.00 | -0.14 | 1.01 |
| Housing Tenure | 1.96 | 48.96 | 1.01 | 1.90 | 47.33 | 1.01 |

The comparison of SOC 2000 and SOC 90 measures of CAMSIS are provided in table 2.60. Log odds and average marginal effects are provided. The CAMSIS models are substantively and statistically identical across all measures. This is unsurprising given their near identical means and standard deviations.

Table 2.81 Comparison of CAMSIS SOC for BCS Model

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | CAMSIS (SOC 2000) | | | Average Marginal Effects | | CAMSIS (SOC 90) | | | Average Marginal Effects | |
| Economic Activity: ‘Don’t Continue Schooling’ Reference Category | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** |
| Continue Schooling |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. |  |  |  |  |  |  |  |  |  |
| *Five or More O’levels* | 1.16 | (0.12) | \*\*\* | 0.25 | (0.02) | 1.17 | (0.12) | \*\*\* | 0.25 | (0.02) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. |  |  |  |  |  |  |  |  |  |
| *Male* | -0.58 | (0.11) | \*\*\* | -0.13 | (0.02) | -0.59 | (0.11) | \*\*\* | -0.13 | (0.02) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. |  |  |  |  |  |  |  |  |  |
| *Do not Own Home* | -0.15 | (0.13) |  | -0.03 | (0.03) | -0.16 | (0.13) |  | -0.03 | (0.03) |
| CAMSIS (SOC 2000) | 0.02 | (0.00) | \*\*\* | 0.00 | (0.00) | 0.02 | (0.00) | \*\*\* | 0.00 | (0.00) |
| Intercept | -0.71 | (0.22) | \*\*\* |  |  | -0.63 | (0.14) | \*\*\* |  |  |
| Number of observations | 1574 | | | | | 1574 | | | | |
| McFadden’s | 0.09 | | | | | 0.09 | | | | |
| McFadden’s Adjusted Pseudo | 0.08 | | | | | 0.08 | | | | |
| Cox-Snell Pseudo | 0.11 | | | | | 0.11 | | | | |
| Nagelkerke Pseudo | 0.16 | | | | | 0.16 | | | | |
| Tjur’s | 0.13 | | | | | 0.13 | | | | |
| AIC | 1936.65 | | | | | 1938.23 | | | | |
| BIC | 1963.46 | | | | | 1965.04 | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: NCDS [Sweeps 0-4]  Note: Sensitivity Analysis of Social Stratification Measures | | | | | | | | | | |

Figure 2.47 details the predictive probabilities at means of economic activity and the average marginal effects for both SOC 2000 and SOC 90 constructions of CAMSIS which once again demonstrates the near identical nature in their construction.

A screenshot of a graph

Description automatically generated

Figure ‎2.46 Comparison of Predictive and AMEs for CAMSIS SOC Codes for BCS Model

Figure 2.48 details all three social stratification measures by SOC 90 construction. Like the NCDS sensitivity analysis, NS-SEC and RGSC SOC 90 models offer the same substantive interpretation whilst CAMSIS offers an alternative substantive picture. Figure 2.49 depicts the coefficients and confidence intervals of all SOC 90 models and compares them directly with the plots of the SOC 2000 counterparts. Looking at figure 2.50 that compares all social stratification measures by their SOC codes also echoes the evidence found from the NCDS based analysis.

A group of black and white squares

Description automatically generated

Figure ‎2.47 Comparison of Predictive and AMEs for all social stratification measures for BCS model for SOC 90 codes

A screenshot of a computer screen

Description automatically generated

Figure ‎2.48 Coefficient plot comparing social stratification measures by SOC

A group of graphs with different lines

Description automatically generated with medium confidence

Figure ‎2.49 Comparison of Predictive and AMEs for all social stratification measures for BCS model for SOC 90 code

#### Discussion and Conclusions

This section attempted to duplicate the analysis of the NCDS cohort by conducting a sensitivity analysis of SOC code constructions for each social stratification measure. The overall findings from the NCDS section are once more repeated here with some caveats. Whilst it is true, like the NCDS analysis that both SOC 90 and SOC 2000 measures of stratification measures produced the same substantive interpretations, unlike the NCDS analysis there was little to no difference in the standard errors between SOC measures. The sensitivity of SOC code constructions found in the NCDS cohort could not be duplicated with the BCS cohort. This could potentially be a factor of time. The BCS cohort were born in 1970, and the dependent variable of interest is set in 1986. This places the BCS cohort very close to both SOC measures. Comparatively the NCDS cohort were born in 1958 and had data analysed from 1974. Perhaps the sensitivity of SOC measures of social class was a factor of members of the NCDS cohort being outside the frame of reference for the SOC codes.

### Handling Missing Data in the BCS

Following the simulation conducted prior to the NCDS analysis, handling missing data methods will be used to identify missingness mechanisms within the BCS cohort. The method of choice is multiple imputation. Across the sweeps of data used within the BCS cohort there are 11,261 individuals that have observations on at least one variable used within analysis. The total complete records analysis for the BCS sample is 1,575. This amounts to a large amount of missingness – 86 per cent. Whilst this substantial amount of missingness at first seems worrying, prior research suggests that so long as the multiple imputation model is appropriately specified a MI model with large amounts of missingness still effectively reduces bias comparative to a complete records based approach (Madley-Dowd *et al.*, 2019; Hyuk Lee and Huber Jr., 2021). To alleviate concerns from conservative users of multiple imputation, a dummy variable adjustment technique will be produced alongside a comparison of a complete records analysis and multiple imputation of the BCS cohort to provide a direct comparison to coefficients and standard errors across each model.

Patterns of missing data are presented in table 2.61. Within the BCS sample, 14 per cent have complete records on all variables, 34 per cent have missing values at the dependent variable of economic activity, 25 per cent have missing values at the dependent variable and educational attainment, a further 10 per cent have missing data exclusively at educational attainment, and finally a further 3 per cent have missing data at NS-SEC[[22]](#footnote-22). In total 1628 cases have a complete observation of all variables.

Table 2.82 Missing data patterns for BCS

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| N | Percent Complete (%) | Educational Attainment | Economic Activity | Housing Tenure | NS-SEC | Sex |
| 1575 | 14 | **✓** | **✓** | **✓** | **✓** | **✓** |
| 3860 | 34 | **✓** |  | **✓** | **✓** | **✓** |
| 2806 | 25 |  |  | **✓** | **✓** | **✓** |
| 1109 | 10 |  | **✓** | **✓** | **✓** | **✓** |
| 387 | 3 | **✓** | **✓** | **✓** |  | **✓** |
| Total = 11,261 |

A total of 11,261 individuals amount to the total eligible sample for the analytical sample covered.

Of the missing proportions of data, the single largest cluster of missingness comes from the dependent variable itself at 34 per cent missingness. This is not surprising, the BCS dataset was not originally envisaged as a dataset to collect economic activity data. Whilst some data was provided this was done so at the 21-sub-sample level where only 10 per cent of the sample were asked, the sweep 5 age 30 dataset where individuals were asked to trace their monthly economic activity since age 16, and again in a special economic activity dataset. This loose connection of datasets has meant that large amounts of missingness exists for several individuals. A second variable of interest is educational attainment which also has a large amount of missingness within the sample. This is primarily due to the way educational data was collected in the BCS. It wasn’t collected at all until age 26. This 10-year gap between sitting examinations and reporting them was further complicated by a massive loss in sample due to a failure to move contact points for the survey from the schools themselves to the individuals place of residence post-mandatory schooling and post-living with parents. Data was once again collected at age 30 of the BCS though like economic activity data there are marginal observations to use.

As addressed previously, a complete records analysis is only considered valid if data can be considered MCAR. Within the descriptive stages of the BCS analysis issues with the data have already been identified – in particular the housing tenure variable appears to provide an inflated level of home owners in comparison to home ownership statistics of the time (HomeOwners Alliance, 2012). Sex is also slightly unrepresentative of the general population and other variables may be affected too. This preliminary evidence suggests a MAR mechanism may be present within this sample. The following section seeks to implement a comparison between complete records analysis and multiple imputation – with the inclusion of the dummy variable adjustment at the educational attainment level as this is the single largest amount of missingness at the independent variable.

When selecting auxiliary variables, following advice from Mostafa and Wiggins (2014) marital status, parity, breastfeeding attempted, mothers age at delivery, mothers age at completion of education, and fathers age at completion of education where selected as auxiliaries for the following multiple imputation model. This provides a total of six auxiliary variables for inclusion in the multiple imputation model, other variables were also suggested but were identical to analytical variables or near identical to them – such as sex or father’s occupation at age 14. For reference the NCDS multiple imputation model had eight auxiliary variables. Trace plots were produced following imputation to make sure means and standard errors were smoothed over iterations.

Following Bodner’s (2008) views on the number of imputations, a dataset such as this with nearly 90 per cent missingness produces peak efficiency with 100+ imputations, 10 imputations are required for 95 per cent confidence in the fractions of missing information and around 258 imputations are required for 95 per cent confidence in the 95 per cent confidence half-widths. Following from the procedure of the NCDS analysis, the largest suggested number of imputations is taken as a baseline and rounded to the nearest 10. This results in 260 imputations for the BCS multiple imputation model. This may sound like many imputations, but it should be remembered that there is no such thing as too many imputations – over-efficiency is not an issue with multiple imputation. Descriptive statistics comparing the complete records analysis and imputation sample are displayed in table 2.62. The multiple imputation model, alongside the complete records analysis and the dummy variable adjustments at educational attainment are presented in table 2.63. Interpretation will follow – both log odds and average marginal effects are provided to aid interpretation.

Table 2.83 Descriptive statistics of BCS sample comparing CRA versus MI

|  |  |  |
| --- | --- | --- |
| Table 2.78: Descriptive Statistics for Economic Activity | | |
|  | CRA% | MI% |
| Continue Schooling or not after September when individuals are 16 |  |  |
| Don't Continue Schooling | 39.64% | 54.47% |
| Continue Schooling | 60.36% | 45.53% |
| Educational Attainment O-levels |  |  |
| <5 O-Levels | 61.05% | 56.11% |
| >5 O-Levels | 38.95% | 43.89% |
| Sex of Respondent |  |  |
| Female | 55.40% | 51.43% |
| Male | 44.60% | 48.57% |
| Housing Tenure of Respondent when Child |  |  |
| Own Home | 76.24% | 64.46% |
| Don't Own Home | 23.76% | 35.54% |
| NS-SEC Social Class of Father when Respondent Child SOC2000 |  |  |
| Large Employers and higher managerial occupations | 5.65% | 4.60% |
| Higher professional occupations | 7.50% | 6.09% |
| Lower Managerial and professional occupations | 18.81% | 15.16% |
| Intermediate occupations | 11.37% | 9.98% |
| Small employers and own account workers | 10.80% | 12.65% |
| Lower supervisory and technical occupations | 16.39% | 15.98% |
| Semi-routine occupations | 12.58% | 15.22% |
| Routine occupations | 16.90% | 20.31% |
|  |  |  |
| n | 1575 | **11261** |
| Data Source: BCS | | |

Table 2.84 Comparison of CRA, Dummy variable adjustment, and MI models for BCS model

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | NS-SEC (SOC200) CRA | | | Average Marginal Effects | | Missingness equal to zero for educational attainment | | | Average Marginal Effects | | Missingness equal to one for educational attainment | | | Average Marginal Effects | | NS-SEC (SOC 2000) MI | | | Average Marginal Effects | |
| Economic Activity: ‘Don’t Continue Schooling’ Reference Category | Coef. | S.E. | Sig. | Prob. | S.E. | Coef. | S.E. | Sig. | Prob. | S.E. | Coef. | S.E. | Sig. | Prob. | S.E. | Coef. | S.E. | Sig. | Prob. | S.E. |
| Continue Schooling |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Less than five O’levels | Ref. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Five or More O’levels | 1.21 | (0.12) | \*\*\* | 0.26 | (0.02) | 1.70 | (0.11) | \*\*\* | 0.37 | (0.02) | -0.25 | (0.09) | \*\* | -0.06 | (0.02) | 1.00 | (0.10) | \*\*\* | 0.22 | (0.01) |
| Sex |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Female | Ref. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Male | -0.58 | (0.11) | \*\*\* | -0.13 | (0.02) | -0.68 | (0.09) | \*\*\* | -0.14 | (0.02) | -0.60 | (0.08) | \*\*\* | -0.14 | (0.02) | -0.64 | (0.08) | \*\*\* | -0.14 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Own Home | Ref. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| Do not Own Home | -0.20 | (0.13) |  | -0.04 | (0.03) | -0.48 | (0.10) | \*\*\* | -0.10 | (0.02) | -0.61 | (0.09) | \*\*\* | -0.14 | (0.02) | -0.56 | (0.09) | \*\*\* | -0.12 | (0.01) |
| NS-SEC |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 1.1 | 0.29 | (0.29) |  | 0.06 | (0.05) | 0.23 | (0.24) |  | 0.05 | (0.05) | 0.19 | (0.22) |  | 0.04 | (0.05) | 0.24 | (0.23) |  | 0.05 | (0.03) |
| 1.2 | 0.12 | (0.26) |  | 0.02 | (0.05) | 0.27 | (0.23) |  | 0.06 | (0.05) | 0.43 | (0.22) | \* | 0.10 | (0.05) | 0.38 | (0.23) |  | 0.08 | (0.03) |
| 2 | Ref. |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| 3 | -0.20 | (0.21) |  | -0.04 | (0.04) | -0.11 | (0.17) |  | -0.02 | (0.04) | -0.15 | (0.16) |  | -0.04 | (0.04) | -0.20 | (0.16) |  | -0.04 | (0.02) |
| 4 | -0.57 | (0.21) | \*\* | -0.12 | (0.05) | -0.58 | (0.17) | \*\*\* | -0.12 | (0.03) | -0.77 | (0.16) | \*\*\* | -0.18 | (0.04) | -0.68 | (0.15) | \*\*\* | -0.15 | (0.02) |
| 5 | -0.70 | (0.19) | \*\*\* | -0.15 | (0.04) | -0.58 | (0.15) | \*\*\* | -0.12 | (0.03) | -0.73 | (0.14) | \*\*\* | -0.17 | (0.03) | -0.70 | (0.14) | \*\*\* | -0.16 | (0.02) |
| 6 | -0.35 | (0.20) |  | -0.07 | (0.04) | -0.33 | (0.16) | \* | -0.07 | (0.03) | -0.61 | (0.15) | \*\*\* | -0.14 | (0.03) | -0.48 | (0.14) | \*\*\* | -0.11 | (0.02) |
| 7 | -0.50 | (0.19) | \*\* | -0.11 | (0.04) | -0.73 | (0.15) | \*\*\* | -0.15 | (0.03) | -1.00 | (0.14) | \*\*\* | -0.24 | (0.03) | -0.88 | (0.14) | \*\*\* | -0.19 | (0.02) |
| Intercept | 0.63 | (0.15) | \*\*\* |  |  | 0.23 | (0.12) | \*\*\* |  |  | 0.91 | (0.12) | \*\*\* |  |  | 0.30 | (0.11) | \*\* |  |  |
| Number of observations | 1575 | | | | | 2684 | | | | | 2684 | | | | | **11261** | | | | |
| Average RVI |  | | | | |  | | | | |  | | | | | 3.19 | | | | |
| Largest FMI |  | | | | |  | | | | |  | | | | | 0.82 | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: BCS [Sweeps 0-5]  Note: CRA, Dummy variable adjustment and MI model for BCS | | | | | | | | | | | | | | | | | | | | |

Interpretation starts with a comparison of the complete records analysis with the dummy variable adjustment models. Dummy variable adjustment is centred on the independent variable with the single largest amount of missingness – educational attainment. There are two dummy variable adjustment models. The first model recodes all missingness at educational attainment equal to zero or ‘less than five O’levels’. The second model recodes all missingness at educational attainment equal to one or ‘five or more O’levels’. Whilst the educational attainment variable itself crosses positive and negative boundaries between the two models and thus makes a substantive interpretation of said variable impossible, a dummy variable adjustment still allows an interpretation of a possible MAR mechanisms and an attempt to identify where that MAR mechanisms originate from. The total n of both models is 2,684. This represents a recapturing of 9 per cent of the total sample.

Both dummy variable adjustment models present identical substantive interpretations with regards to sex comparative to the complete records analysis – average marginal effects are within 1 per cent of the CRA. Housing tenure for both dummy variable adjustment models is statistically significant and substantively different from the CRA model. Both report a decreased log odds of continuing schooling for individuals that grew up in homes not owned by their parents compared to those that do. This translates into a 10-14 per cent decreased probability in terms of average marginal effect. This finding is suggestive of a MAR mechanism being present – the original CRA underplays the effect of housing tenure within the BCS cohort. The change in statistical significance on top of the overrepresentation of homeowners in the BCS sample also suggests that non-response from non-homeowners is part of the cause of the MAR mechanism within the CRA.

The NS-SEC variable is consistent across dummy variable adjustment models and the CRA model except for NS-SEC 7 and NS-SEC 6. Both dummy variable models find NS-SEC 6 to be statistically significant, translated into average marginal effects the models present a range of 7-14 per cent decreased probability of continuing schooling compared to NS-SEC 2 peers. Whilst all three models interpret NS-SEC 7 as having a decreased log odds of continuing schooling compared to reference category NS-SEC 2, the average marginal effects present a slightly different substantive interpretation. Whilst the CRA model translates this effect into a 11 per cent decreased probability to continue schooling, the dummy variable adjustment models have average marginal effects between 15-24 per cent. Whilst not as large a difference as the housing tenure variable, this does indicate that some of the MAR mechanism may be identified within NS-SEC 7 social origins.

Whilst the dummy variable adjustment models have provided some benefit in establishing possible MAR mechanisms with a comparison to the CRA there are some issues with solely using this approach as detailed in Jones (1996), a more concerning issue presents itself in the lack of sample recovery from this particular dummy variable adjustment. Whilst 9 per cent of the sample is recovered, this is still below 50 per cent of the sample in the BCS. Whilst MAR mechanisms have been identified there is no evidence to suggest that with greater sample recovery this could not reverse or increase in size. This is therefore why multiple imputation is such an important tool and will now be used in comparison to the findings provided thus far.

Starting with the educational attainment variable, the MI model shows a decrease in the overall log odds reducing the average marginal effects from 26 per cent in the CRA to 22 per cent in the MI model. Sex remains consistent between the CRA and MI models – with a 1 per cent difference not ultimately impacting the substantive interpretation. Housing tenure provides a substantive effect that is closer to both dummy adjustment models than the CRA model. This confirms the finding that the MAR mechanism is at least partially identified within the housing tenure variable – of individuals that grew up in homes not owned by their parents. The variable NS-SEC is mostly consistent between the CRA and the MI models except for NS-SEC 4 and NS-SEC 7. NS-SEC 4 sees a 3 per cent average marginal effect difference between the two models and NS-SEC 7 has a 7 per cent difference. This again is suggestive of a MAR mechanism being present primarily for individuals of social origins NS-SEC 4 and 7.

The coefficients and confidence intervals for both CRA and MI models are presented in figure 2.51 to aid interpretation. The figure demonstrates the reduced confidence intervals that a MI model produces. It also notes the most substantively significant divergences between the two models. Housing tenure’s coefficient within the MI model is below that of the lower confidence interval of the CRA model for example – the same is true for NS-SEC 7. This once more presents evidence that this is where the MAR mechanism is most prominent. Predicted probabilities and average marginal effects between the two models for NS-SEC is also presented in figure 2.52.

A graph with red and blue lines

Description automatically generated

Figure ‎2.50 Comparison coefficient plot of CRA and MI models

A screenshot of a computer screen

Description automatically generated

Figure ‎2.51 Comparison of CRA and MI NS-SEC margins

#### Discussion and Conclusions

This section had one primary aim. To investigate missingness within the BCS cohort based on the analytical models presented in this chapter. After conducting appropriate handling missing data methods via multiple imputation, a comparison of a multiple imputation model on BCS data and a complete records analysis was conducted. The results suggest that the BCS sample used to analyse youth transitions has evidence of a MAR mechanism present. On further inspection this MAR mechanism seems to originate from date focused upon housing tenure and select categories within NS-SEC. This evidence is provided from the multiple imputation model that provides alternative substantive interpretations of the analytical model proposed for the BCS sample. Given these substantive differences, it is appropriate to focus primarily upon the substantive interpretations of the multiple imputation model going forward.

This does present some difficulties, however. Given the original pooled dataset analysis at the start of Part 1, and the findings in this section that the BCS sample has evidence of a MAR mechanism, the original interpretation of the pooled analysis also comes into question. Without this handling missing data investigation, this finding would have been lost and an inadequate interpretation of the analysis would have occurred. The relative size of the difference in substantive interpretation between complete records analysis and multiple imputation model demonstrate the need of handling missing data methods within social scientific research. Moving forward, a re-analysis of the pooled analysis will need to be conducted to correct for the bias encountered within the complete records analysis sample of the BCS dataset. This re-analysis will require a conditional multiple imputation of the BCS sample of the pooled datasets.

### Discussion and Conclusions

This chapter focused upon the British Cohort Study. The BCS was analysed in a granular format duplicating the analysis present for the NCDS cohort. The intent behind this duplicate granular analysis was an attempt to understand the within cohort differences – through social stratification and SOC code sensitivity analysis, as well as identifying any handling missing data issues present within the individual samples. The findings from the social stratification sensitivity analysis and the SOC code sensitivity analysis for the BCS cohort was remarkably similar to that of the NCDS sample. The conclusions thus drawn from the BCS analysis need not bear repeating once more, the NCDS conclusions on these matters serve as the conclusions for the BCS sample also.

The BCS sample diverges in one major instance from the NCDS sample. That is on the matter of handling missing data. Whilst the NCDS data provided a healthy sample with no evidence of a MAR mechanism, the BCS sample had many variable related difficulties that resulted in a relatively small analytical sample compared to the overall sample size of the BCS waves. This in turn presented evidence of a MAR mechanism within the analytical sample. The amount of missingness present in the BCS sample stretches contemporary handling missing data methods to their methodological limit – this meant that other less than gold standard handling missing data methods were also employed to provide some extra rigidity to the analysis of missing data within the BCS.

Results for the BCS cohort indicate that structural influences still hold influence upon individuals’ youth transitions. The socio-historical context in which the youth of the BCS cohort live presents pressures on young people continuing schooling depending upon their educational attainment, sex, housing tenure, and social class origins. The comparative difference between the BCS and other cohorts requires re-examination following evidence of a MAR mechanism. Before this, it is inappropriate to draw strong conclusions in a comparative context.

## Granular UKHLS Analysis

### Sensitivity Analysis of Social Stratification Measures using UKHLS

#### Testing Measures of Parental Social Class

#### Discussion and Conclusion

### SOC Code Sensitivity analysis using UKHLS

#### Measuring SOC Codes

#### Discussion and Conclusions

### Handling Missing Data in the UKHLS

#### Discussion and Conclusions

### Discussion and Conclusions

## A Return to Modelling First Transition

All three cohorts have now been analysed in granular detail, exploring sensitivity analyses of social stratification measures, SOC codes, and handling missing data procedures. The latter of this granular analysis has identified that the BCS cohort has a MAR mechanism within its CRA sample. On further investigation this is producing bias estimates and erroneous substantive interpretation of the real effects of given analytical variables within the proposed analytical model. This not only is producing bias estimates for the BCS cohort but, in the original analysis of combined cohorts, producing bias estimates for the entire analysis. This section is a return to modelling first transitions across all cohorts used for analysis. Conditional multiple imputation is used to re-create and compare the original modelling first transition analysis. Firstly, the CRA model will be produced to echo table 2.62 with logs odds and average marginal effects provided. Secondly, a direct comparison will be made with the final model of interest and interpretation – a model of combined cohorts, with a cohort analytical variable to understand cohort specific effects, with each analytical variable having an interaction with said cohort effect. This CRA model is compared directly to its MI counterpart using log odds and average marginal effects to assess the level of bias present from a lack of imputation of the BCS cohort. Other cohorts are not imputed due to investigation finding no evidence for a MAR mechanism. The table is present in table 2.64.

The comparison of CRA and MI models for the BCS have substantively been made in prior sections. Therefore, the substantive interpretation here will focus primarily upon the cohort level differences accounting for the CRA and MI models. For members of the BCS cohort, men see a decrease in the average probability of continuing schooling from 8 to 11 per cent. The cohort specific effect decreases from 16 to 12 per cent resulting in an overall effect of a 1 per cent increased probability for men to continue schooling in the BCS compared to the NCDS. Housing tenure for the BCS nearly triples in terms of log odds, with members that grew up in homes not owned by their parents represents a 10 per cent decreased probability for the MI model – though when adding the cohort specific effect this results in a 2 per cent increased probability to continue schooling compared to the NCDS counterparts.

NS-SEC 4, 5, and 7 all see an increase in the effect size for BCS members. The largest single increase occurred for NS-SEC 7 members of the BCS, where there was a 7 per cent decreased probability of continuing schooling in the CRA model compared to a 16 per cent decreased probability in the MI model. Even with the cohort specific effect added, for NS-SEC 4, 5 and 7 members still have a decreased probability of continuing schooling compared to the NCDS peers.

Table 2.85 Descriptive Statistics of First Destinations

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Table 1: Descriptive Statistics for Economic Activity | | | | |
|  | CRA=n | CRA% | MI=n | MI% |
| Continue Schooling or not after September when individuals are 16 |  |  |  |  |
| Don't Continue Schooling |  | 41.05% |  | 45.36% |
| Continue Schooling |  | 58.95% |  | 54.65% |
| Educational Attainment O-levels |  |  |  |  |
| <5 O-Levels |  | 63.97% |  | 58.18% |
| >5 O-Levels |  | 36.03% |  | 41.82% |
| Sex of Respondent |  |  |  |  |
| Female |  | 50.95% |  | 50.87% |
| Male |  | 49.05% |  | 49.13% |
| Housing Tenure of Respondent when Child |  |  |  |  |
| Own Home |  | 52.53% |  | 57.18% |
| Don't Own Home |  | 47.47% |  | 42.82% |
| NS-SEC Social Class of Father when Respondent Child SOC2000 |  |  |  |  |
| Large Employers and higher managerial occupations |  | 3.51% |  | 3.91% |
| Higher professional occupations |  | 5.29% |  | 5.57% |
| Lower Managerial and professional occupations |  | 13.36% |  | 13.93% |
| Intermediate occupations |  | 9.85% |  | 9.83% |
| Small employers and own account workers |  | 11.96% |  | 12.40% |
| Lower supervisory and technical occupations |  | 16.32% |  | 16.11% |
| Semi-routine occupations |  | 16.86% |  | 16.32% |
| Routine occupations |  | 22.85% |  | 21.92% |
| Cohort |  |  |  |  |
| NCDS |  | 84.24% |  | 42.76% |
| BCS |  | 15.76% |  | 57.24% |
|  |  |  |  |  |
| n |  | 9985 |  | **19672** |
| Data Source: BCS | | | | |

Table 2.86 Descriptive Statistics of First Destinations by Cohort

|  |  |  |  |
| --- | --- | --- | --- |
| Descriptive Statistics by Cohort | | | |
|  | Member of Cohort | | |
|  | NCDS | BCS | Total |
| % |  |  |  |
| Economic Activity of individual September when 16 |  |  |  |
| Don’t Continue Schooling | 41.31% | 56.09% | 45.36% |
| Continue Schooling | 58.69% | 43.91% | 54.65% |
| Educational Attainment O'levels |  |  |  |
| <5 O-Levels | 64.51% | 49.85% | 58.18% |
| >5 O-Levels | 35.49% | 50.15% | 41.82% |
| Sex of Respondent |  |  |  |
| Female | 50.11% | 51.43% | 50.87% |
| Male | 49.89% | 48.57% | 49.13% |
| Housing Tenure of Respondent when Child |  |  |  |
| Own Home | 48.09% | 64.45% | 57.46% |
| Don't Own Home | 51.91% | 35.55% | 42.54% |
| NS-SEC Social Class of Father when Respondent Child SOC2000 |  |  |  |
| 1.1 | 3.10% | 4.58% | 3.95% |
| 1.2 | 4.87% | 6.09% | 5.57% |
| 2 | 12.34% | 15.16% | 13.96% |
| 3 | 9.57% | 9.97% | 9.80% |
| 4 | 12.17% | 12.65% | 12.45% |
| 5 | 16.31% | 16.02% | 16.14% |
| 6 | 17.66% | 15.21% | 16.25% |
| 7 | 23.97% | 20.32% | 21.88% |

Table 2.87 Modelling First Major Transition with Combined Cohorts (Imputed Models)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Pooled Model | | | Average Marginal Effects | | Imputation Model | | | Average Marginal Effects | |
| Economic Activity: ‘Don’t Continue Schooling’ Reference Category | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** |
| Continue Schooling |  |  |  |  |  |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |  |  |  |  |  |
| *Less than five O’levels* | Ref. |  |  |  |  |  |  |  |  |  |
| *Five or More O’levels # NCDS* | 2.98 | (0.06) | \*\*\* | 0.55 | (0.01) | 2.98 | (0.06) | \*\*\* | 0.52 | (0.01) |
| *Five or More O’levels # BCS* | 1.21 | (0.12) | \*\*\* | 0.19 | (0.02) | 1.00 | (0.10) | \*\*\* | 0.19 | (0.02) |
| Sex |  |  |  |  |  |  |  |  |  |  |
| *Female* | Ref. |  |  |  |  |  |  |  |  |  |
| *Male # NCDS* | -0.50 | (0.06) | \*\*\* | -0.07 | (0.01) | -0.50 | (0.06) | \*\*\* | -0.09 | (0.01) |
| *Male # BCS* | -0.58 | (0.11) | \*\*\* | -0.08 | (0.01) | -0.64 | (0.08) | \*\*\* | -0.11 | (0.01) |
| Housing Tenure |  |  |  |  |  |  |  |  |  |  |
| *Own Home* | Ref. |  |  |  |  |  |  |  |  |  |
| *Do not Own Home # NCDS* | -0.63 | (0.06) | \*\*\* | -0.09 | (0.01) | -0.63 | (0.06) | \*\*\* | -0.11 | (0.01) |
| *Do not Own Home # BCS* | -0.20 | (0.13) |  | -0.03 | (0.02) | -0.56 | (0.09) | \*\*\* | -0.10 | (0.02) |
| NS-SEC (SOC 2000) |  |  |  |  |  |  |  |  |  |  |
| *1.1 # NCDS* | 0.02 | (0.19) |  | 0.00 | (0.03) | 0.02 | (0.19) |  | 0.00 | (0.04) |
| *1.1 # BCS* | 0.29 | (0.29) |  | 0.04 | (0.04) | 0.24 | (0.23) |  | 0.04 | (0.04) |
| *1.2 # NCDS* | 0.48 | (0.17) | \*\* | 0.08 | (0.03) | 0.48 | (0.17) | \*\* | 0.09 | (0.03) |
| *1.2 # BCS* | 0.12 | (0.26) |  | 0.02 | (0.04) | 0.38 | (0.23) |  | 0.07 | (0.04) |
| *2* | Ref. |  |  |  |  |  |  |  |  |  |
| *3 # NCDS* | -0.25 | (0.13) | \* | -0.04 | (0.02) | -0.25 | (0.13) | \* | -0.05 | (0.02) |
| *3 # BCS* | -0.20 | (0.21) |  | -0.03 | (0.03) | -0.20 | (0.16) |  | -0.04 | (0.03) |
| *4 # NCDS* | -0.89 | (0.12) | \*\*\* | -0.14 | (0.02) | -0.89 | (0.12) | \*\*\* | -0.16 | (0.02) |
| *4 # BCS* | -0.57 | (0.21) | \*\* | -0.08 | (0.03) | -0.68 | (0.16) | \*\*\* | -0.12 | (0.03) |
| *5 # NCDS* | -0.76 | (0.11) | \*\*\* | -0.12 | (0.02) | -0.76 | (0.11) | \*\*\* | -0.13 | (0.02) |
| *5 # BCS* | -0.70 | (0.19) | \*\*\* | -0.09 | (0.02) | -0.70 | (0.14) | \*\*\* | -0.13 | (0.03) |
| *6 # NCDS* | -0.89 | (0.11) | \*\*\* | -0.14 | (0.02) | -0.89 | (0.11) | \*\*\* | -0.15 | (0.02) |
| *6 # BCS* | -0.35 | (0.20) |  | -0.05 | (0.03) | -0.48 | (0.15) | \*\* | -0.09 | (0.03) |
| *7 # NCDS* | -1.11 | (0.11) | \*\*\* | -0.17 | (0.02) | -1.11 | (0.11) | \*\*\* | -0.19 | (0.02) |
| *7 # BCS* | -0.50 | (0.19) | \*\* | -0.07 | (0.03) | -0.88 | (0.14) | \*\*\* | -0.16 | (0.02) |
| Cohort |  |  |  |  |  |  |  |  |  |  |
| *NCDS* | Ref. |  |  |  |  |  |  |  |  |  |
| *BCS* | 1.03 | (0.18) | \*\*\* | 0.16 | (0.03) | 0.71 | (0.15) | \*\*\* | 0.12 | (0.03) |
| Intercept | -0.40 | (0.09) | \*\*\* |  |  | -0.40 | (0.09) | \*\*\* |  |  |
| Number of Observations | 9985 | | | | | 19672 | | | | |
| McFadden’s | 0.35 | | | | | Average RVI: 1.59 | | | | |
| McFadden’s Adjusted Pseudo | 0.34 | | | | | Largest FMI: 0.82 | | | | |
| Cox-Snell Pseudo | 0.38 | | | | |  | | | | |
| Nagelkerke Pseudo | 0.51 | | | | |  | | | | |
| Tjur’s | 0.42 | | | | |  | | | | |
| AIC | 8954.76 | | | | |  | | | | |
| BIC | 9113.35 | | | | |  | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: NCDS [Sweeps 0-4] & BCS [Sweeps 0-5]  Note: Each model following the ‘cohort model’ adds an interaction variable with each analytical variable and the cohort variable | | | | | | | | | | |

Average marginal effects and predicted margins are provided in graphical form from figures 2.51-2.54. Starting with NS-SEC, figure 2.51 presents the average marginal effects and predicted probabilities for both the CRA and MI models. The general comparison point to be made is the relative stability across cohorts that the MI model demonstrates. The CRA model produces a larger amount of cross-cohort variation within NS-SEC categories compared to the MI model.

A graph with lines and dots

Description automatically generated

A screenshot of a graph

Description automatically generated

Figure ‎2.52 Comparison of Pooled Average Marginal Effects and Predicted Probabilities of NS-SEC by Cohort (CRA on left, MI on right)

Figure 2.52 depicts the predicted probabilities of sex across CRA and MI models. The substantive difference across the two models demonstrates the relative strength educational attainment has within the NCDS cohort in terms of continuing schooling. The MI model sees an even greater reduction in that strength for the BCS cohort. Whilst both cohorts share a similar overall trend – that individuals with five or more O’levels will have a higher predicted probability of continuing schooling – this trend is much stronger for the NCDS cohort than the BCS cohort. Even stronger than first though under the CRA model. Whilst there are changes between the CRA and MI models, substantively speaking the interpretation of patterns within the data remains the same.

A screenshot of a computer screen

Description automatically generated

Figure ‎2.53 Comparison of Pooled Average Marginal Effects and Predicted Probabilities of Educational Attainment by Cohort (CRA on left, MI on right)

Figure 2.53 depicts the predicted probabilities of sex by cohort comparing the CRA and MI models. The difference between CRA and MI models demonstrates that the MI model sees virtually identical results for men and women between the NCDS and BCS cohorts. The CRA and MI models present substantively the same results – though the MI model does increase the strength of association.

A screenshot of a computer screen

Description automatically generated

Figure ‎2.54 Comparison of Pooled Average Marginal Effects and Predicted Probabilities of Sex by Cohort (CRA on left, MI on right)

Figure 2.54 depicts the predicted probabilities of housing tenure by cohort for a comparison of CRA and MI models. Whereas the CRA models suggest that the NCDS cohort has a general steep slope of continuing schooling dependent on housing tenure, the BCS cohort has no such effect. The MI model by comparison sees the BCS and NCDS cohorts having a near identical effect. The comparison between CRA and MI models changes the substantive interpretation of the analytical model.

A screenshot of a computer screen

Description automatically generated

Figure ‎2.55 Comparison of Pooled Average Marginal Effects and Predicted Probabilities of Housing Tenure by Cohort (CRA on left, MI on right)

### Discussion and Conclusions

This section saw a re-analysis of previously examined pooled data. This was required due to in-depth investigations of individual cohorts presenting evidence of a MAR mechanism within the BCS sample cohort. This re-analysis of pooled data thus performed a conditional imputation on the BCS sample of the pooled data. Whilst providing a substantial contribution to the literature on the sociology of youth, this section also serves as a direct example of why handling missing data methods are crucial to understand social phenomena. The comparison of complete records analysis and multiple imputation models identified key variables within the analytical models where the substantive interpretation of patterns within the data differed. Specifically housing tenure and NS-SEC sub-categories were identified to produce bias estimates within the BCS cohort that left unchecked would have been interpreted as a non-substantive to weak effect on continuing schooling. Employing effective handling missing data procedures correctly identified these bias estimators under a missing at random mechanism and were reinterpreted using an updated conditionally imputed model. This multiple imputation model identifies housing tenure and NS-SEC as key predictors in the impact of continuing schooling for both the NCDS and BCS cohorts.

For both cohorts, educational attainment was the single largest impact on continuing schooling. Though the strength of educational attainment on youths first transition appears to be deteriorating at a rather rapid pace from cohort to cohort. Other structural inequalities have maintained a constant impact upon youths first transitions, with sex and housing tenure not substantively changing across cohorts. This provides evidence for a persistent structural effect for both variables. It also provides substantive evidence against theories of New Structuralism – housing tenure does not appear to be becoming the dominant form of stratification. In addition to this, it also does not appear accurate to report that other traditional stratification measures such as sex are deteriorating – in direct contradiction of theories of Individualisation. There is, however, a minor degradation in the impact of social class on youths first transition, though it must be recognising this change is small. Across most NS-SEC categories, the strength of their impact on youths first transition appears to degrade by a few percentage points between cohorts. Finally, it is worth mentioning that the nature of when a person is born presents a very real impact on youths first transition. The cohort effect whilst not the largest single effect within this analytical model, is substantial. The socio-historical context that someone is born into, grows up in, and transitions from youth to adulthood in has a very real impact on youths first transition.

This section has compared the complete records model with the multiple imputation model to identify any differences within the substantive interpretations of either model. Certain differences were identified. Not performing adequate handling missing data methods would have resulted in presenting biased estimators and interpreting incorrect overall effects from the proposed analytical model. Overall structural inequalities have a substantive impact on a youths first transition – the strength of that inequality depends based upon the type of inequality and the time that inequality exists within.

## Discussion and Conclusions for Part 1

Part 1 of this thesis centred upon a youths first transition, dichotomised by one of the first real choices a young person is given on their journey to adulthood. The decision to continue schooling or not after they reach the school leaving age is an important sociological phenomenon to investigate. The choice made at this life-stage has impacts that reflect an individual’s wider life course. Part 1 has dealt with pooled cohort data, attempting to understand not only structural inequalities varies impacts upon youths first transition, but also how different cohorts in varied socio-historical contexts operate under differing conditions. Part 1 started with a pooled analysis and interpretation of these various cohorts. Each cohort was then singled out and investigated in greater depth. This was done to assess the sensitivity of different cohorts to different social stratification measure of social class and social distance – to see if different manifestations of stratification change the overall substantive conclusions of an analytical model dependent upon the socio-historical context of that model. In a similar fashion, a sensitivity analysis of those stratification measures using different standard occupational codes was also conducted. This was to assess the time sensitivity of using different SOC constructions of the same stratification variables. Finally, each cohort was investigated using handling missing data methods. This was performed to assess the risk of a missing at random mechanism and to provide solutions to any biased estimators being presented and interpretated. The findings from each of these sections will now be discussed:

A sensitivity analysis of social stratification measures comparing: NS-SEC, RGSC, and CAMSIS measures was conducted across each cohort sample. The findings showed remarkable consistency across all cohort samples. The overall findings were NS-SEC and RGSC based models reported virtually identical substantive findings across cohort samples. The only distinct quality between the two measures related to the manual/non-manual sensitivity of the RGSC measure. This was made particularly salient in the older cohort samples like the NCDS but became less relevant through each successive cohort. The NS-SEC measure still captured this divide where it was present, but due to RGSCs construction it was much more evident within its analytical models. Finally with respect to CAMSIS as a measure of social stratification, models using CAMSIS across cohort samples came to identical conclusions. CAMSIS as a variable presented statistically significant results with near zero substantive significance across all samples. The conclusion drawn from these findings is that a measure of social distance is not particularly relevant to an analytical strategy that seeks to understand youths first transitions. Given the nature of a social distance scale, it would perhaps be of greater importance for inclusion in models of later transitions – though that is out of scope for this thesis.

A sensitivity of standard occupational classification codes followed. This analysis had some cross-cohort variability in its findings though general conclusions are shared by all cohort samples. The NCDS cohort found that using SOC 2000 codes compared to SOC 90 codes provided for overall smaller standard errors across the model. The BCS cohort also found this but to a much less extent that the NCDS cohort sample. The primary driver behind this appears to be in the construction of the SOC codes themselves. SOC 90 codes and SOC 2000 codes primarily differ in the reorganisation and addition of occupations related to the technology, engineering, and finance fields – occupations that simply did not exist in the 1990s (and certainly not in 1958). In addition to these new occupations, however, was a reorganisation of traditional occupational categories. This reorganisation appears to be the result of smaller standard errors across analytical models using NS-SEC and RGSC SOC 2000 constructions. This provides some debate into which SOC construction one should use when using historic data such as the NCDS. Whilst smaller standard errors present a data-driven, statistically orientated reason for using a SOC 2000 construction the socio-historical reality of the occupational makeup of 1974 (when NCDS youth make their first transition) is more accurately reflected in the SOC 90 construction of either NS-SEC or RGSC than their SOC 2000 counterparts. If these were the only arguments regarding SOC measures, then following a sociologically relevant approach, the SOC 90 measure would be selected for analysis of older cohorts. However, considering the comparative nature of this thesis demands a level of consistency between sample cohort analyses. In this instance, using a SOC 2000 construction across all cohort samples makes more sense than using a SOC 90 construction for one sample, and SOC 2000 construction for the other cohort samples. This section has sought to present reasonable interpretation of two different constructions of social stratification variables and attempt to present arguments for and against use of certain SOC constructions dependent on needs.

Finally, handling missing data procedures were conducted across each sample cohort to assess evidence of a missing at random mechanism and thus biased estimators. Prior to any methods being employed on the cohorts themselves, an investigation into several handling missing data methods was performed on simulated data. This simulation intended to bring clarity to existing arguments on handling missing data methods – with relation gold standard (and non-gold standard) approaches, as well as add to the literature by performing direct comparisons on the effectiveness of Full Information Maximum Likelihood and Multplie Imputation based approaches. The simulation found that there is not much different between the two methods – though the simulation favoured multiple imputation, though only slightly. Following this multiple imputation was performed on all cohort samples. The NCDS sample identified no MAR mechanism present, whilst the BCS sample did identify a MAR mechanism and bias estimators. Following this identification, the BCS complete records analysis was directly compared with the multiple imputed models and re-interpreted. The results show that some variables in the BCS sample were indeed biased. This section provided direct comparisons of handling missing data procedures and implemented these strategies within the thesis cohort samples.

The finding of bias estimators in the BCS sample required a re-analysis of pooled data. Due to a MAR mechanism only being found present in one of the cohorts samples a conditional imputation was conducted, keeping all other samples as their complete records versions whilst conducting multiple imputation on the BCS sample and pooling cohorts together. A direct comparison between the pooled complete records analysis and the imputed pooled analysis again displayed amongst a few identified variables biased estimators and were thus re-interpreted using the imputed pooled analysis.

Part 1 of this thesis conducted a large number of statistical tests and interpretation of youths first transition. Reflecting to the story of youths first transition across cohorts some concluding thoughts are established. A key portion of this thesis was outlining the various theories of structure and agency and their interpretations of youth sociology. Whilst theories of Individualisation stated that there was a marked decline in structural inequalities influence on human action, and New Structuralisms thesis that new social cleavages would overtake traditional social inequalities impacts on human behaviour; part 1 of this thesis provides categorical proof negating both these theories’ claims. Structural inequalities have for the most part remained markedly static – sex and housing tenure especially. The only variable in the analytical models that presents a marked decline is educational attainment. This particularly variable is susceptible to socio-historical context related to grade inflation, the rise of credentialism, and the exogenous shocks of a collapsing youth labour market pushing young people into post-16 choices that they are constrained to make. The marked static nature of housing tenure is a key argument against the thesis of New Structuralism – new social cleavages do not appear to be becoming the dominant form of stratification in British society – in fact traditional social cleavages remain strong – such as sex and social class.

Individuals first transition is characterised by social constraint on part by structural forces related educational attainment, sex, housing tenure, and social class. These structural factors are somewhat sensitivity to the socio-historical context in which individuals are born into. Indeed, the cohort specific effects noted within the analytical models make this point clear – socio-historical context matters for youths first transitions. The life course of young people entering their school-to-work transition for the first time is impacted by structural forces as well as being sensitivity to the socio-historical context of the time. These findings seem congruent with theories of structuralism, life course, and structured individualism.

# Youth’s First Destination post-mandatory schooling

## First Destination

The analysis up until this point has dealt with a binary indicator of youth first transitions from school-to-work. This indicator has been dichotomised through ‘Continuing Schooling’ versus ‘Not Continuing Schooling’. Whilst this presents a direct comparison in the evolution of youths first transition to continuing schooling over the course of cohort change, it is ultimately an overly simplistic characterisation of the multitude of pathways that young people encountered during their school-to-work transitions. By dichotomising youth transitions in this way, there is a loss of information surrounding other important youth destinations, such as apprenticeship programmes, employment, and unemployment & out of the labour force (OLF). The following chapter uses the same analytical sample from the previous chapters, using a conditionally imputed sample of the BCS based on handling missing data investigations to assess the role structural inequalities have upon youths first destinations. This chapter will briefly discuss the descriptive picture of this sample – primarily focusing on a new ‘economic activity’ dependent variable that considers four distinct first destination opportunities following mandatory schooling. These are: Continue schooling, employment, apprenticeships, unemployment & OLF. Following this descriptive breakdown, an analytical model is presented using log odds and average marginal effects with substantive interpretation. This chapter will conclude by assessing the potential new information gained from exploring first destinations over first transitions of the following chapters.

As mentioned, the dependent variable, ‘economic activity’ has four categories. The following analytical model will use multinominal logistic regression using economic activity as the dependent variable and the same independent variables outlined in prior chapters. The descriptive picture of the independent viables has not changed, and for all purposes is the identical sample to the sample used in all other chapters. The only difference is the operationalisation of the dependent variable that has been transformed from a binary variable – of which a logistic regression model was used – to a nominal variable – of which a multinominal logistic regression model will be used.

Descriptive statistics are provided in table 3.1. Only the economic activity variable will be reflected upon as all independent variables have been discussed in previous chapters. Economic activity has four categories. The first is employment with 36.18 per cent of the sample. Schooling has the plurality of the sample located in its category at 42.51 per cent. Apprenticeships is the third largest category with 18.47 per cent. Unemployment & OLF has the lowest proportion of the sample with 2.83 per cent.

Table 3.1 Descriptive Statistics for Youth's First Destinations

|  |  |  |
| --- | --- | --- |
|  | Descriptive Statistics for Economic Activity | |
|  | n | % |
| Economic Activity of individual September when 16 |  |  |
| Employment | 1,781,451 | 34.75% |
| Schooling | 2,189,533 | 42.71% |
| Apprenticeship | 993,490 | 19.38% |
| Unemployment & OLF | 161,827 | 3.16% |
| Educational Attainment O'levels |  |  |
| <5 O-Levels | 3,065,827 | 59.77% |
| >5 O-Levels | 2,063,691 | 40.23% |
| Sex of Respondent |  |  |
| Female | 2,611,687 | 50.87% |
| Male | 2,522,704 | 49.13% |
| Housing Tenure of Respondent when Child |  |  |
| Own Home | 2,949,510 | 57.46% |
| Don't Own Home | 2,184,023 | 42.54% |
| NS-SEC Social Class of Father when Respondent Child SOC2000 |  |  |
| 1.1 | 202,713 | 3.95% |
| 1.2 | 285,868 | 5.57% |
| 2 | 716,272 | 13.96% |
| 3 | 503,009 | 9.80% |
| 4 | 638,756 | 12.45% |
| 5 | 828,584 | 16.14% |
| 6 | 834,192 | 16.25% |
| 7 | 1,123,112 | 21.88% |
| Member of Cohort |  |  |
| NCDS | 2,195,271 | 42.76% |
| BCS | 2,939,121 | 57.24% |
|  |  |  |
|  |  | 100% |
| Data Source: NCDS & BCS  Note: Conditional Imputed Sample on BCS Cohort | | |

Table 3.2 provides a richer breakdown of the dependent variable by Cohort. This table provides detailed comparisons of the descriptive differences in youth’s first destinations by cohort. For the NCDS cohort there is a near identical split between individuals entering employment at 38.25 per cent and schooling at 39.17 per cent. This trend does not continue into the BCS cohort, with schooling making up the majority of the cohort at 60.36 per cent

Table 3.2 Descriptive Statistics for Youth's First Destinations by Cohort

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Descriptive Statistics by Economic Activity | | | | | |
|  | Member of Cohort | | | | |
| NCDS | | BCS | | Total |
| n | % | n | % |  |
| Economic Activity of individual September when 16 |  |  |  |  |  |
| Employment | 839,637 | 38.25% | 941,814 | 32.13% | 34.75% |
| Schooling | 859,995 | 39.17% | 1,329,538 | 45.36% | 42.71% |
| Apprenticeship | 428,301 | 19.51% | 565,189 | 19.28% | 19.38% |
| Unemployment & OLF | 67,338 | 3.07% | 94,489 | 3.22% | 3.16% |
| Educational Attainment O'levels |  |  |  |  |  |
| <5 O-Levels | 1,416,186 | 64.51% | 1,649,641 | 56.22% | 59.77% |
| >5 O-Levels | 779,085 | 35.49% | 1,284,606 | 43.78% | 40.23% |
| Sex of Respondent |  |  |  |  |  |
| Female | 1,100,115 | 50.11% | 1,511,572 | 51.43% | 50.87% |
| Male | 1,095,156 | 49.89% | 1,427,548 | 48.57% | 49.13% |
| Housing Tenure of Respondent when Child |  |  |  |  |  |
| Own Home | 1,055,745 | 48.09% | 1,893,765 | 64.45% | 57.46% |
| Don't Own Home | 1,139,526 | 51.91% | 1,044,497 | 35.55% | 42.54% |
| NS-SEC Social Class of Father when Respondent Child SOC2000 |  |  |  |  |  |
| 1.1 | 68,121 | 3.10% | 134,592 | 4.58% | 3.95% |
| 1.2 | 107,010 | 4.87% | 178,858 | 6.09% | 5.57% |
| 2 | 270,918 | 12.34% | 445,354 | 15.16% | 13.96% |
| 3 | 210,105 | 9.57% | 292,904 | 9.97% | 9.80% |
| 4 | 267,264 | 12.17% | 371,492 | 12.65% | 12.45% |
| 5 | 358,092 | 16.31% | 470,492 | 16.02% | 16.14% |
| 6 | 387,585 | 17.66% | 446,607 | 15.21% | 16.25% |
| 7 | 526,176 | 23.97% | 596,936 | 20.32% | 21.88% |

Interpretation of the model presented in table 3.3 will follow…

Table 3.3 Multinominal Logistic Regression model of conditionally imputed pooled dataset investigating youths first destination

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Imputation Model | | | Average Marginal Effects | |
| Economic Activity: ‘Continue Schooling’ Reference Category | **Coef.** | **S.E.** | **Sig.** | **Prob.** | **S.E.** |
| Employment |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  |  |  |
| *Five or More O’levels # NCDS* | -3.07 | (0.08) | \*\*\* | -0.33 | (0.01) |
| *Five or More O’levels # BCS* | -0.79 | (0.12) | \*\*\* | -0.05 | (0.02) |
| Sex |  |  |  |  |  |
| *Female* | Ref. | (.) |  |  |  |
| *Male # NCDS* | -0.11 | (0.07) |  | -0.17 | (0.01) |
| *Male # BCS* | 0.72 | (0.09) | \*\*\* | 0.10 | (0.02) |
| Housing Tenure |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  |  |  |
| *Do not Own Home # NCDS* | 0.72 | (0.07) | \*\*\* | 0.10 | (0.01) |
| *Do not Own Home # BCS* | 0.57 | (0.09) | \*\*\* | 0.06 | (0.02) |
| NS-SEC (SOC 2000) |  |  |  |  |  |
| *1.1 # NCDS* | 0.07 | (0.22) |  | 0.03 | (0.04) |
| *1.1 # BCS* | -0.28 | (0.27) |  | -0.04 | (0.05) |
| *1.2 # NCDS* | -0.53 | (0.21) | \* | -0.06 | (0.03) |
| *1.2 # BCS* | -0.36 | (0.25) |  | -0.05 | (0.05) |
| *2* | Ref. | (.) |  |  |  |
| *3 # NCDS* | 0.33 | (0.14) | \* | 0.06 | (0.03) |
| *3 # BCS* | 0.11 | (0.19) |  | -0.01 | (0.04) |
| *4 # NCDS* | 0.90 | (0.13) | \*\*\* | 0.10 | (0.02) |
| *4 # BCS* | 0.59 | (0.19) | \*\* | 0.04 | (0.04) |
| *5 # NCDS* | 0.78 | (0.12) | \*\*\* | 0.09 | (0.02) |
| *5 # BCS* | 0.65 | (0.17) | \*\*\* | 0.05 | (0.03) |
| *6 # NCDS* | 0.92 | (0.12) | \*\*\* | 0.11 | (0.02) |
| *6 # BCS* | 0.41 | (0.17) | \* | 0.03 | (0.03) |
| *7 # NCDS* | 1.20 | (0.12) | \*\*\* | 0.16 | (0.02) |
| *7 # BCS* | 0.76 | (0.16) | \*\*\* | 0.05 | (0.03) |
| Cohort |  |  |  |  |  |
| *NCDS* | Ref. | (.) |  |  |  |
| *BCS* | -0.99 | (0.17) | \*\*\* | -0.19 | (0.03) |
| Intercept | 0.12 | (0.10) |  |  |  |
| Apprenticeship |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  |  |  |
| *Five or More O’levels # NCDS* | -2.79 | (0.09) | \*\*\* | -0.16 | (0.01) |
| *Five or More O’levels # BCS* | -1.32 | (0.18) | \*\*\* | -0.11 | (0.02) |
| Sex |  |  |  |  |  |
| *Female* | Ref. | (.) |  |  |  |
| *Male # NCDS* | 1.81 | (0.08) | \*\*\* | 0.32 | (0.01) |
| *Male # BCS* | 0.53 | (0.10) | \*\*\* | 0.02 | (0.01) |
| Housing Tenure |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  |  |  |
| *Do not Own Home # NCDS* | 0.42 | (0.08) | \*\*\* | -0.01 | (0.01) |
| *Do not Own Home # BCS* | 0.55 | (0.11) | \*\*\* | 0.03 | (0.02) |
| NS-SEC (SOC 2000) |  |  |  |  |  |
| *1.1 # NCDS* | -0.13 | (0.25) |  | -0.02 | (0.03) |
| *1.1 # BCS* | -0.23 | (0.33) |  | -0.01 | (0.04) |
| *1.2 # NCDS* | -0.34 | (0.23) |  | -0.01 | (0.03) |
| *1.2 # BCS* | -0.47 | (0.35) |  | -0.04 | (0.04) |
| *2* | Ref. | (.) |  |  |  |
| *3 # NCDS* | 0.17 | (0.17) |  | 0.00 | (0.02) |
| *3 # BCS* | 0.35 | (0.24) |  | 0.04 | (0.03) |
| *4 # NCDS* | 0.92 | (0.15) | \*\*\* | 0.06 | (0.02) |
| *4 # BCS* | 0.81 | (0.21) | \*\*\* | 0.07 | (0.03) |
| *5 # NCDS* | 0.82 | (0.14) | \*\*\* | 0.06 | (0.02) |
| *5 # BCS* | 0.76 | (0.20) | \*\*\* | 0.05 | (0.03) |
| *6 # NCDS* | 0.85 | (0.14) | \*\*\* | 0.04 | (0.02) |
| *6 # BCS* | 0.55 | (0.21) | \*\* | 0.04 | (0.03) |
| *7 # NCDS* | 0.93 | (0.14) | \*\*\* | 0.03 | (0.02) |
| *7 # BCS* | 1.00 | (0.19) | \*\*\* | 0.08 | (0.03) |
| Cohort |  |  |  |  |  |
| *NCDS* | Ref. | (.) |  |  |  |
| *BCS* | 0.21 | (0.21) |  | 0.10 | (0.02) |
| Intercept | -1.48 | (0.13) | \*\*\* |  |  |
| Unemployment & OLF |  |  |  |  |  |
| Educational Attainment |  |  |  |  |  |
| *Less than five O’levels* | Ref. | (.) |  |  |  |
| *Five or More O’levels # NCDS* | -3.43 | (0.26) | \*\*\* | -0.03 | (0.00) |
| *Five or More O’levels # BCS* | -2.85 | (1.01) | \*\* | -0.04 | (0.01) |
| Sex |  |  |  |  |  |
| *Female* | Ref. | (.) |  |  |  |
| *Male # NCDS* | -0.07 | (0.14) |  | -0.02 | (0.00) |
| *Male # BCS* | 0.47 | (0.21) | \* | -0.00 | (0.00) |
| Housing Tenure |  |  |  |  |  |
| *Own Home* | Ref. | (.) |  |  |  |
| *Do not Own Home # NCDS* | 0.92 | (0.16) | \*\*\* | 0.02 | (0.01) |
| *Do not Own Home # BCS* | 0.70 | (0.24) | \*\* | 0.01 | (0.01) |
| NS-SEC (SOC 2000) |  |  |  |  |  |
| *1.1 # NCDS* | -0.36 | (0.64) |  | -0.01 | (0.01) |
| *1.1 # BCS* | -0.22 | (1.21) |  | 0.01 | (0.03) |
| *1.2 # NCDS* | -1.93 | (1.03) |  | -0.03 | (0.01) |
| *1.2 # BCS* | 0.40 | (0.85) |  | 0.03 | (0.04) |
| *2* | Ref. | (.) |  |  |  |
| *3 # NCDS* | -0.31 | (0.38) |  | -0.01 | (0.01) |
| *3 # BCS* | 0.16 | (0.80) |  | 0.00 | (0.02) |
| *4 # NCDS* | 0.55 | (0.31) |  | -0.00 | (0.01) |
| *4 # BCS* | 1.04 | (0.63) |  | 0.02 | (0.02) |
| *5 # NCDS* | 0.20 | (0.30) |  | -0.01 | (0.01) |
| *5 # BCS* | 1.37 | (0.56) | \* | 0.03 | (0.02) |
| *6 # NCDS* | 0.78 | (0.27) | \*\* | 0.00 | (0.01) |
| *6 # BCS* | 0.71 | (0.62) |  | 0.01 | (0.02) |
| *7 # NCDS* | 1.03 | (0.26) | \*\*\* | 0.01 | (0.01) |
| *7 # BCS* | 1.46 | (0.55) | \*\* | 0.03 | (0.02) |
| Cohort |  |  |  |  |  |
| *NCDS* | Ref. | (.) |  |  |  |
| *BCS* | -1.04 | (0.57) |  | -0.02 | (0.02) |
| Intercept | -2.26 | (0.25) | \*\*\* |  |  |
| Number of Observations | 19672 | | | | |
| Average RVI | 1.81 | | | | |
| Largest FMI | 0.92 | | | | |
| \*\*\* p<.001, \*\* p<.01, \* p<.05 Data Source: NCDS [Sweeps 0-4] & BCS [Sweeps 0-5]  Note: Conditionally imputed at the BCS cohort | | | | | |

A graph of different colored lines

Description automatically generated with medium confidence

Figure ‎3.1 Predictive and average marginal effects of NS-SEC on youth's first destination by cohorts

A screenshot of a computer

Description automatically generated

Figure ‎3.2 Predictive and average marginal effects of educational attainment on youth's first destination by cohorts

A screenshot of a computer

Description automatically generated

Figure ‎3.3 Predictive and average marginal effects of sex on youth's first destination by cohorts

A screenshot of a computer

Description automatically generated

Figure ‎3.4 Predictive and average marginal effects of housing tenure on youth's first destination by cohorts

A screenshot of a computer screen

Description automatically generated

Figure ‎3.5 Coefficient plots of each category of youth's first destination

### Discussion and Conclusions

## Discussion and Conclusions for Part 2

# Conclusions

## Introduction to Part 5

## Substantive Conclusions

## Methodological Reflections

## Final Remarks

# 

# Appendix

A graph with lines and dots

Description automatically generated

Figure ‎5.1 Coefficient Plot of RGSC model

A diagram of a graph

Description automatically generated

Figure ‎5.2 log odds versus quasi-variance statistics of RGSC for NCDS model

A graph with numbers and lines

Description automatically generated with medium confidence

Figure ‎5.3 Coefficient Plot of CAMSIS model

A screenshot of a computer screen

Description automatically generated

Figure ‎5.4 Predictive and AMEs of RGSC for NCDS model

A graph of a line and a line

Description automatically generated with medium confidence

Figure ‎5.5 Predictive and AMEs of CAMSIS for NCDS model

A graph with red and black lines

Description automatically generated

Figure ‎5.6 log odds versus quasi-variance statistics for NS-SEC SOC 90

A diagram of a graph

Description automatically generated

Figure ‎5.7 log odds versus quasi-variance statistics of RGSC SOC 90 for NCDS Model

A graph showing different colored lines

Description automatically generated with medium confidence

Figure ‎5.8 Trace plot summaries for Economic Activity

A graph showing different colored lines

Description automatically generated

Figure ‎5.9 Trace plot summaries for Educational Attainment

A graph showing different colored lines

Description automatically generated with medium confidence

Figure ‎5.10 Trace plot summaries for NS-SEC

A graph showing the number of numbers

Description automatically generated with medium confidence

Figure ‎5.11 Trace plot summaries for Housing Tenure

A diagram with red and black lines

Description automatically generated

Figure ‎5.12 Log odds versus Quasi-Variance Statistics for BCS model (RGSC)

A screenshot of a graph

Description automatically generated

Figure ‎5.13 Predictive and AMEs of RGSC for BCS model

A screenshot of a graph

Description automatically generated

Figure ‎5.14 Predictive and AMEs of CAMSIS for BCS model

A diagram with red and black lines

Description automatically generated

Figure ‎5.15 log odds versus quasi-variance statistics for NS-SEC SOC 90 (BCS model)

A graph with red and black lines

Description automatically generated

Figure ‎5.16 log odds versus quasi-variance statistics for RGSC SOC 90 (BCS model)

# Bibliography

Data Acknowledgements

Gregg, P., 2012, *Occupational Coding for the National Child Development Study (1969, 1991-2008) and the 1970 British Cohort Study (1980, 2000-2008)*, [data collection], University of London, Institute of Education, Centre for Longitudinal Studies, University of London, Institute of Education, Centre for Longitudinal Studies, [original data producer(s)]. Accessed 22 February 2024. SN: 7023, [DOI: http://doi.org/10.5255/UKDA-SN-7023-1](http://doi.org/10.5255/UKDA-SN-7023-1)

University of London, Institute of Education, Centre for Longitudinal Studies, 2023, *National Child Development Study: Childhood Data from Birth to Age 16, Sweeps 0-3, 1958-1974*, [data collection], National Children's Bureau, *3rd Edition,*National Children's Bureau, National Birthday Trust Fund, [original data producer(s)]. Accessed 22 February 2024. SN: 5565, [DOI: http://doi.org/10.5255/UKDA-SN-5565-2](http://doi.org/10.5255/UKDA-SN-5565-2)

University of London, Institute of Education, Centre for Longitudinal Studies, 2023, *National Child Development Study: Age 23, Sweep 4, 1981, and Public Examination Results, 1978*, [data collection], National Children's Bureau, *2nd Edition,*National Children's Bureau, [original data producer(s)]. Accessed 22 February 2024. SN: 5566, [DOI: http://doi.org/10.5255/UKDA-SN-5566-1](http://doi.org/10.5255/UKDA-SN-5566-1)

Chamberlain, G., Chamberlain, R., University of London, Institute of Education, Centre for Longitudinal Studies, 2023, *1970 British Cohort Study: Birth and 22-Month Subsample, 1970-1972*, [data collection], UK Data Service, *3rd Edition,*Accessed 22 February 2024. SN: 2666, [DOI: http://doi.org/10.5255/UKDA-SN-2666-2](http://doi.org/10.5255/UKDA-SN-2666-2)

University of London, Institute of Education, Centre for Longitudinal Studies, Butler, N., Bynner, J., 2023, *1970 British Cohort Study: Age 10, Sweep 3, 1980*, [data collection], UK Data Service, *7th Edition,*Accessed 22 February 2024. SN: 3723, [DOI: http://doi.org/10.5255/UKDA-SN-3723-8](http://doi.org/10.5255/UKDA-SN-3723-8)

Bynner, J., Butler, N., University of London, Institute of Education, Centre for Longitudinal Studies, 2023, *1970 British Cohort Study: Age 16, Sweep 4, 1986*, [data collection], UK Data Service, *9th Edition,*Accessed 22 February 2024. SN: 3535, [DOI: http://doi.org/10.5255/UKDA-SN-3535-6](http://doi.org/10.5255/UKDA-SN-3535-6)

Bynner, J., 2023, *1970 British Cohort Study: Age 21 Sample Survey, 1992*, [data collection], UK Data Service, *3rd Edition,*Accessed 22 February 2024. SN: 4715, [DOI: http://doi.org/10.5255/UKDA-SN-4715-2](http://doi.org/10.5255/UKDA-SN-4715-2)

University of London, Institute of Education, Centre for Longitudinal Studies, Bynner, J., 2023, *1970 British Cohort Study: Age 26, Sweep 5, 1996*, [data collection], UK Data Service, *5th Edition,*Accessed 22 February 2024. SN: 3833, [DOI: http://doi.org/10.5255/UKDA-SN-3833-3](http://doi.org/10.5255/UKDA-SN-3833-3)

University of London, Institute of Education, Centre for Longitudinal Studies, 2023, *1970 British Cohort Study: Age 29, Sweep 6, 1999-2000*, [data collection], Joint Centre for Longitudinal Research, *4th Edition,*Joint Centre for Longitudinal Research, [original data producer(s)]. Accessed 22 February 2024. SN: 5558, [DOI: http://doi.org/10.5255/UKDA-SN-5558-3](http://doi.org/10.5255/UKDA-SN-5558-3)

University College London, UCL Institute of Education, Centre for Longitudinal Studies, 2023, *1970 British Cohort Study: Activity Histories, 1986-2016*, [data collection], UK Data Service, *4th Edition,*Accessed 22 February 2024. SN: 6943, [DOI: http://doi.org/10.5255/UKDA-SN-6943-4](http://doi.org/10.5255/UKDA-SN-6943-4)

University of Essex, Institute for Social and Economic Research. (2023). *Understanding Society: Waves 1-13, 2009-2022 and Harmonised BHPS: Waves 1-18, 1991-2009: Special Licence Access*. [data collection]. *17th Edition.*UK Data Service. SN: 6931, [DOI: http://doi.org/10.5255/UKDA-SN-6931-16](http://doi.org/10.5255/UKDA-SN-6931-16)

Department for Education, University of Essex, Institute for Social and Economic Research. (2024). *Understanding Society: Linked Education Administrative Datasets (National Pupil Database), England, 1995-2018: Secure Access*. [data collection]. *3rd Edition.*UK Data Service. SN: 7642, [DOI: http://doi.org/10.5255/UKDA-SN-7642-3](http://doi.org/10.5255/UKDA-SN-7642-3)

Reference List

‘A New Training Initiative: A Programme for Action’ (1981) *Government White Paper* [Preprint].

Akaike, H. (1998) ‘Information Theory and an Extension of the Maximum Likelihood Principle’, in E. Parzen, K. Tanabe, and G. Kitagawa (eds) *Selected Papers of Hirotugu Akaike*. New York, NY: Springer New York (Springer Series in Statistics), pp. 199–213. Available at: https://doi.org/10.1007/978-1-4612-1694-0\_15.

Alcott, B. (2013) ‘Predicting departure from British education: Identifying those most at risk through discrete time hazard modelling’, *Widening Participation and Lifelong Learning*, 15(4), pp. 46–64. Available at: https://doi.org/10.5456/WPLL.15.4.46.

Allison, P. (2012a) ‘Handling Missing Data by Maximum Likelihood’, *SAS Global Forum* [Preprint].

Allison, P. (2012b) ‘Why Maximum Likelihood is Better Than Multiple Imputation’, *Statistical Horizons*, 9 July. Available at: https://statisticalhorizons.com/ml-better-than-mi/ (Accessed: 15 May 2023).

Allison, P. (2013) ‘What’s the Best R-Squared for Logistic Regression?’, *Statistical Horizons*, 13 February. Available at: https://statisticalhorizons.com/r2logistic/ (Accessed: 22 February 2024).

Allison, P. (2015) ‘Maximum Likelihood is Better than Multiple Imputation: Part II’, *Statistical Horizons*, 5 May. Available at: https://statisticalhorizons.com/ml-is-better-than-mi/ (Accessed: 15 May 2023).

Anders, J. and Dorsett, R. (2017) ‘What young English people do once they reach school-leaving age: A cross-cohort comparison for the last 30 years’, *Longitudinal and Life Course Studies*, 8(1). Available at: https://doi.org/10.14301/llcs.v8i1.399.

Archer, M.S. (1995) *Realist social theory: The morphogenetic approach*. Cambridge University Press.

Arnett, J.J. (2000) ‘Emerging adulthood: A theory of development from the late teens through the twenties.’, *American Psychologist*, 55(5), pp. 469–480. Available at: https://doi.org/10.1037/0003-066X.55.5.469.

Arnett, J.J. (2006) ‘Emerging Adulthood in Europe: A Response to Bynner’, *Journal of Youth Studies*, 9(1), pp. 111–123. Available at: https://doi.org/10.1080/13676260500523671.

Arulampalam, W. and Booth, A.L. (1997) ‘Who gets over the training hurdle? A study of the training experiences of young men and women in Britain’, *Journal of Population Economics*, 10(2), pp. 197–217. Available at: https://doi.org/10.1007/s001480050038.

Arulampalam, W.N. and Booth, A.L. (2001) ‘Learning and Earning: Do Multiple Training Events Pay? A Decade of Evidence from a Cohort of Young British Men’, *Economica*, 68(271), pp. 379–400. Available at: https://doi.org/10.1111/1468-0335.00252.

Ashton, D., Maguire, M. and Garland, V. (1982) ‘Youth in the Labour Market’, *Department of Employment*, 34.

Ashton, D., Maguire, M. and Spilsbury, M. (2016) *Restructuring the labour market: The implications for youth.* Springer.

Baudrillard, J. (1988) *Selected Writings*. Mark Poster.

*BBC* (1982) ‘1982: UK unemployment tops three million’, 26 January. Available at: http://news.bbc.co.uk/onthisday/hi/dates/stories/january/26/newsid\_2506000/2506335.stm (Accessed: 11 June 2024).

Beck, U. (1992) *Risk Society: Towards a new modernity*. SAGE.

Beck, U. (2002) *Individualisation: Institutionalized Individualism and its Social and Political Consequences*. SAGE Publications.

Beck, U. (2014) *The brave new world of work*. John Wiley & Sons.

Beck, U., Giddens, A. and Lash, S. (1994) *Reflexive modernization: Politics, tradition and aesthetics in the modern social order*. Stanford University Press.

Bell, D.N.F. and Blanchflower, D.G. (2013) ‘How to Measure Underemployment?’

Bergman, M.M. and Joye, D. (2001) ‘Comparing Social Stratification Schemas: CAMSIS, CSP-CH, Goldthorpe, ISCO-88, Treiman, and Wright’, *Cambridge studies in Social research*, p. 53.

Bernardi, L., Huinink, J. and Settersten, R.A. (2019) ‘The life course cube: A tool for studying lives’, *Advances in Life Course Research*, 41, p. 100258. Available at: https://doi.org/10.1016/j.alcr.2018.11.004.

Birnbaum, N. (2002) *After Progress*. Oxford University Press. Available at: https://doi.org/10.1093/acprof:oso/9780195158595.001.0001.

Blanchflower, D. and Lynch, L. (1992) *Training at Work: A Comparison of U.S. and British Youths*. w4037. Cambridge, MA: National Bureau of Economic Research, p. w4037. Available at: https://doi.org/10.3386/w4037.

Bland, R. (1979) ‘Measuring Social Class’, *Sociology*, 13, pp. 283–91.

Blanden, J. and Machin, S. (2017) ‘Home Ownership and Social Mobility’, *CEP Discussion Paper* [Preprint].

Blanden, J. and Macmillan, L. (2014) ‘Education and Intergenerational Mobility: Help or Hindrance?’, *Centre for Analysis of Social Exclusion* [Preprint].

Blundell, R. *et al.* (2000) ‘The Returns to Higher Education in Britain: Evidence From a British Cohort’, *The Economic Journal*, 110(461), pp. F82–F99. Available at: https://doi.org/10.1111/1468-0297.00508.

Blundell, R., Dearden, L. and Sianesi, B. (2001) ‘Estimating the Returns to Education: Models, Methods and Results’, *Centre for the Economics of Education* [Preprint].

Bodner, T.E. (2008) ‘What Improves with Increased Missing Data Imputations?’, *Structural Equation Modeling: A Multidisciplinary Journal*, 15(4), pp. 651–675. Available at: https://doi.org/10.1080/10705510802339072.

Boero, G. *et al.* (2020) ‘HOW DOES THE RETURN TO A DEGREE VARY BY CLASS OF AWARD?’, *Higher Education Statistics Agency* [Preprint].

Booth, A.L. and Satchell, S.E. (1994) ‘APPRENTICESHIPS AND JOB TENURE’, *Oxford Economic Papers*, 46(4), pp. 676–695. Available at: https://doi.org/10.1093/oxfordjournals.oep.a042153.

Bottero, W. (2004) ‘Class Identities and the Identity of Class’, *Sociology*, 38(5), pp. 985–1003. Available at: https://doi.org/10.1177/0038038504047182.

Breen, R. (2022) ‘The stubborn persistence of educational inequality’, *IFS Deaton Review* [Preprint].

Breen, R., Karlson, K.B. and Holm, A. (2013) ‘Total, Direct, and Indirect Effects in Logit and Probit Models’, *Sociological Methods & Research*, 42(2), pp. 164–191. Available at: https://doi.org/10.1177/0049124113494572.

Brooks, R. (2009) *Transitions from education to work: new perspectives from Europe and beyond*. Springer.

Brückner, H. and Mayer, K.U. (2005) ‘De-Standardization of the Life Course: What it Might Mean? And if it Means Anything, Whether it Actually Took Place?’, *Advances in Life Course Research*, 9, pp. 27–53. Available at: https://doi.org/10.1016/S1040-2608(04)09002-1.

Brunello, G. and Rocco, L. (2017) ‘The Labor Market Effects of Academic and Vocational Education over the Life Cycle: Evidence Based on a British Cohort’, *Journal of Human Capital*, 11(1), pp. 106–166. Available at: https://doi.org/10.1086/690234.

Buck, N. and McFall, S. (2011) ‘Understanding Society: design overview’, *Longitudinal and Life Course Studies* [Preprint].

Bukodi, E. (2009) ‘Education, First Occupation and Later Occupational Attainment: Cross-cohort Changes among Men and Women in Britain’, *CLS Cohort Studies*, 4.

Bukodi, E., Bourne, M. and Betthäuser, B. (2017) ‘Wastage of talent?’, *Advances in Life Course Research*, 34, pp. 34–42. Available at: https://doi.org/10.1016/j.alcr.2017.09.003.

Bukodi, E. and Dex, S. (2010) ‘Bad Start: Is There a Way Up? Gender Differences in the Effect of Initial Occupation on Early Career Mobility in Britain’, *European Sociological Review*, 26(4), pp. 431–446. Available at: https://doi.org/10.1093/esr/jcp030.

Bukodi, E. and Goldthorpe, J.H. (2009) ‘Class Origins, Education and Occupational Attainment: Cross-cohort Changes among Men in Britain’, *CLS Cohort Studies*, 3.

Bukodi, E. and Goldthorpe, J.H. (2011) ‘Social class returns to higher education: chances of access to the professional and managerial salariat for men in three British birth cohorts’, *Longitudinal and Life Course Studies*, 2(2). Available at: https://doi.org/10.14301/llcs.v2i2.122.

Bukodi, E., Goldthorpe, J.H. and Kuha, J. (2017) ‘The pattern of social fluidity within the British class structure: a topological model’, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 180(3), pp. 841–862. Available at: https://doi.org/10.1111/rssa.12234.

Bynner, J. (1998) ‘Education and Family Components of Identity in the Transition from School to Work’, *International Journal of Behavioral Development*, 22(1), pp. 29–53. Available at: https://doi.org/10.1080/016502598384504.

Bynner, J. (1999) ‘New Routes to Employment: Integration and Exclusion’, in W.R. Heinz (ed.) *From Education to Work*. 1st edn. Cambridge University Press, pp. 65–86. Available at: https://doi.org/10.1017/CBO9780511527876.005.

Bynner, J. *et al.* (2002) ‘Young people’s changing routes to independence’, *Joseph Rowntree Foundation* [Preprint].

Bynner, J. (2005) ‘Rethinking the Youth Phase of the Life-course: The Case for Emerging Adulthood?’, *Journal of Youth Studies*, 8(4), pp. 367–384. Available at: https://doi.org/10.1080/13676260500431628.

Bynner, J. (2012) ‘Policy Reflections Guided by Longitudinal Study, Youth Training, Social Exclusion, and More Recently Neet’, *British Journal of Educational Studies*, 60(1), pp. 39–52. Available at: https://doi.org/10.1080/00071005.2011.650943.

Bynner, J. (2017) ‘1970 British Cohort Study (BCS70) Twenty one-year Sample Survey’, *CLS Cohort Studies* [Preprint].

Bynner, J. (2023) ‘1970 British Cohort Study: Age 21 Sample Survey, 1992’. UK Data Service. Available at: https://doi.org/10.5255/UKDA-SN-4715-2.

Bynner, J., Butler, N. and University College London (2023) ‘1970 British Cohort Study: Age 16, Sweep 4, 1986’. Institute of Education, Centre for Longitudinal Studies. Available at: https://doi.org/10.5255/UKDA-SN-3535-6.

Bynner, J. and Ferri, E. (2003) *Changing Britain, Changing Lives*. Institute of Education Press.

Bynner, J., Ferri, E. and Shepherd, P. (2019) *Twenty-something in the 1990s: Getting on, getting by, getting nowhere*. Routledge.

Bynner, J. and Joshi, H. (2002) ‘Equality and Opportunity in Education: Evidence from the 1958 and 1970 birth cohort studies’, *Oxford Review of Education*, 28(4), pp. 405–425. Available at: https://doi.org/10.1080/0305498022000013599.

Bynner, J. and Parsons, S. (2000) ‘Marginalization and Value Shifts under the Changing Economic Circumstances Surrounding the Transition to Work: A Comparison of Cohorts Born in 1958 and 1970’, *Journal of Youth Studies*, 3(3), pp. 237–249. Available at: https://doi.org/10.1080/713684379.

Bynner, J., Wiggins, R. and Parsons, S. (1996) ‘AN EXPLORATORY COMPARATIVE ANALYSIS OF DATA COLLECTED IN THE 1958 AND 1970 BRITISH BIRTH COHORT STUDIES’:, *Conference of the International Sociological Association* [Preprint].

Canaan, S. *et al.* (2022) ‘Maternity Leave and Paternity Leave: Evidence on the Economic Impact of Legislative Changes in High Income Countries’.

Carpenter, J.R. and Kenward, M. (2012) *Multiple imputation and its application*. John Wiley & Sons.

*Carr Committee Report on Recruitment and Training in Industry* (1958). Available at: https://mrc-catalogue.warwick.ac.uk/records/NUT/4/1/3/54 (Accessed: 5 July 2024).

Cavanaugh, J.E. and Neath, A.A. (2019) ‘The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements’, *WIREs Computational Statistics*, 11(3), p. e1460. Available at: https://doi.org/10.1002/wics.1460.

Cebulla, A. and Tomaszewski, W. (2013) ‘The demise of certainty: shifts in aspirations and achievement at the turn of the century’, *International Journal of Adolescence and Youth*, 18(3), pp. 141–157. Available at: https://doi.org/10.1080/02673843.2013.767743.

Chamberlain, G., Chamberlain, R. and University of London (2023) ‘1970 British Cohort Study: Birth and 22-Month Subsample, 1970-1972’. Institute of Education, Centre for Longitudinal Studies. Available at: https://doi.org/10.5255/UKDA-SN-2666-2.

Chen, A. (2014) ‘QV: Stata module to compute quasi-variances’, *Statistical Software Components* [Preprint].

Chevalier, A. and Lanot, G. (2001) ‘The Relative Effect of Family and Financial Characteristics on Educational Achievement’, p. 32.

Clarke, L. (1978) ‘The Transition for School to Work: A critical review of literature’, *Report*, 48.

Cockburn, C. (1987) *Two-track training: Sex inequalities and the YTS.* Macmillan Education.

Collins, L.M., Schafer, J.L. and Kam, C.-M. (2001) ‘A comparison of inclusive and restrictive strategies in modern missing data procedures.’, *Psychological Methods*, 6(4), pp. 330–351. Available at: https://doi.org/10.1037/1082-989X.6.4.330.

Conlon, G. (2001) *The differential in earnings premia between academically and vocationally trained males in the United Kingdom*. London: Centre for the Economics of Education.

Connelly, R., Gayle, V. and Lambert (2016) ‘Statistical modelling of key variables in social survey data analysis’, *Methodological Innovations*, 9, p. 205979911663800. Available at: https://doi.org/10.1177/2059799116638002.

Connelly, R., Gayle, V. and Lambert, Paul S. (2016) ‘A review of educational attainment measures for social survey research’, *Methodological Innovations*, 9. Available at: https://doi.org/10.1177/2059799116638001.

Connelly, R., Gayle, V. and Lambert, Paul S (2016) ‘Ethnicity and ethnic group measures in social survey research’, *Methodological Innovations*, 9, p. 205979911664288. Available at: https://doi.org/10.1177/2059799116642885.

Connolly, S. and Gregory, M. (2010) ‘Dual tracks: part-time work in life-cycle employment for British women’, *Journal of Population Economics*, 23(3), pp. 907–931. Available at: https://doi.org/10.1007/s00148-009-0249-4.

Connolly, S., Micklewright, J. and Nickell, S. (1992) ‘THE OCCUPATIONAL SUCCESS OF YOUNG MEN WHO LEFT SCHOOL AT SIXTEEN \*’, *Oxford Economic Papers*, 44(3), pp. 460–479. Available at: https://doi.org/10.1093/oxfordjournals.oep.a042058.

Cox, D.R. and Snell, E.J. (1989) *Analysis of Binary Data.* Second Edition. Chapman Hall: New York.

*Crowther Report Volume I* (1959). Available at: http://www.educationengland.org.uk/documents/crowther/crowther1959-1.html (Accessed: 10 November 2022).

Croxford, L. *et al.* (2006) ‘TRENDS IN EDUCATION AND YOUTH TRANSITIONS ACROSS BRITAIN 1984-2002’, *Conference: Education and Social Change: England, Wales and Scotland 1984-2002*, p. 22.

Davis, K. and Moore, W. (1994) ‘The functions of stratification’, in *Social Stratification in Sociological Perspectives*. Westview Press.

Deakin, B.M. (1996) *The youth labour market in Britain: the role of intervention (No. 62).* Cambridge University Press.

Devine, F. (2017) ‘The “new structuralism”: class politics and class analysis’, in *Social Class and Marxism*. Taylor & Francis.

Dex, S. and Bukodi, E. (2012) ‘The Effects of Part-Time Work on Women’s Occupational Mobility in Britain: Evidence from the 1958 Birth Cohort Study’, *National Institute Economic Review*, 222, pp. R20–R37. Available at: https://doi.org/10.1177/002795011222200103.

Dex, S., Ward, K. and Joshi, H. (2008) ‘Gender differences in occupational wage mobility in the 1958 cohort’, *Work, Employment and Society*, 22(2), pp. 263–280. Available at: https://doi.org/10.1177/0950017008089104.

Di Salvo, P. and Ermisch, J. (1997) ‘Analysis of the Dynamics of Housing Tenure Choice in Britain’, *Journal of Urban Economics*, 42(1), pp. 1–17. Available at: https://doi.org/10.1006/juec.1996.2009.

Diewald, M. and Mayer, K.U. (2008) ‘The sociology of the life course and life span psychology: integrated paradigm or complementing pathways?’, p. 24.

Dodgeon, B. (2002) ‘Longitudinal Linkage in BCS70: Rationalising Case Identifiers’, *CLS Cohort Studies* [Preprint].

Dolton, P., Galinda-Rueda, F. and Makepeace, G. (2004) ‘The Long Term Effects of Government Sponsored Training’, 20.

Dolton, P., Joshi, H. and Makepeace, G. (2002) ‘Unpacking Unequal Pay Between Men and Women Across Cohort and Lifecycle’, *CLS Cohort Studies* [Preprint].

Dolton, P., Makepeace, G. and Marcenaro‐Gutierrez, O.D. (2005) ‘Career progression: Getting‐on, getting‐by and going nowhere’, *Education Economics*, 13(2), pp. 237–255. Available at: https://doi.org/10.1080/09645290500031447.

Dorsett, R. and Lucchino, P. (2013) ‘Visualising the school-to-work transition: an analysis using optimal matching’, *Journal of Social Policy* [Preprint].

Droy, L., Goodwin, J. and O’connor, H. (2019) ‘Liminality, Marginalisation and Low-Skilled Work: Mapping long-term labour market difficulty following participation in the 1980s government-sponsored youth training schemes (YTS)’, *Occasional Papers* [Preprint]. Available at: https://doi.org/10.13140/RG.2.2.28494.92486.

Duckworth, K. and Schoon, I. (2012) ‘Beating the Odds: Exploring the Impact of Social Risk on Young People’s School-to-Work Transitions during Recession in the UK’, *National Institute Economic Review*, 222, pp. R38–R51. Available at: https://doi.org/10.1177/002795011222200104.

Dustmann, C. *et al.* (1996) ‘Earning and Learning: Educational Policy and the Growth of Part-Time Wurk by Full-Time Pupils’, *The journal of applied public economics* [Preprint].

Duta, A. and Iannelli, C. (2018) ‘Social Class Inequalities in Graduates’ Labour Market Outcomes: The Role of Spatial Job Opportunities’, *Social Sciences*, 7(10), p. 201. Available at: https://doi.org/10.3390/socsci7100201.

Duta, A., Iannelli, C. and Breen, R. (2021) ‘Social inequalities in attaining higher education in Scotland: New evidence from sibling data’, *British Educational Research Journal*, 47(5), pp. 1281–1302. Available at: https://doi.org/10.1002/berj.3725.

Duta, A., Wielgoszewska, B. and Iannelli, C. (2020) *Different degrees of career success: Social origin and graduates’ education and labour market trajectories*. Available at: https://doi.org/10.1016/j.alcr.2020.100376.

Elder, G. (1995) ‘Life trajectories in changing societies’, in Zimmerman, B. J., *Self-Efficacy in Changing Societies*. 1st edn. Edited by A. Bandura. Cambridge University Press, pp. 202–231. Available at: https://doi.org/10.1017/CBO9780511527692.009.

Elder, G., Johnson, M. and Crosnoe, R. (2003) ‘The emergence and development of life course theory’, in *Handbook of the Lifecourse*. Springer.

Elder, G.H. (1994) ‘Time, Human Agency, and Social Change: Perspectives on the Life Course’, *Social Psychology Quarterly*, 57(1), p. 4. Available at: https://doi.org/10.2307/2786971.

Elliott, J. and Shepherd, P. (2006) ‘Cohort Profile: 1970 British Birth Cohort (BCS70)’, *International Journal of Epidemiology*, 35(4), pp. 836–843. Available at: https://doi.org/10.1093/ije/dyl174.

Enders, C.K. (2001) ‘A Primer on Maximum Likelihood Algorithms Available for Use With Missing Data’, *Structural Equation Modeling: A Multidisciplinary Journal*, 8(1), pp. 128–141. Available at: https://doi.org/10.1207/S15328007SEM0801\_7.

Enders, C.K. (2010) *Applied missing data analysis*. New York: Guilford Press (Methodology in the social sciences).

Erikson, R., Goldthorpe, J.H. and Portocarero, L. (1979) ‘Intergenerational Class Mobility in Three Western European Societies: England, France and Sweden’, *The British Journal of Sociology*, 30(4), p. 415. Available at: https://doi.org/10.2307/589632.

Erikson, R., Goldthorpe, J.H. and Portocarero, L. (1982) ‘Social Fluidity in Industrial Nations: England, France and Sweden’, *The British Journal of Sociology*, 33(1), p. 1. Available at: https://doi.org/10.2307/589335.

Erikson, R., Goldthorpe, J.H. and Portocarero, L. (1983) ‘Intergenerational Class Mobility and the Convergence Thesis: England, France and Sweden’, *Brit. Jnl. of Sociology* [Preprint].

Evans, G. (1992) ‘Testing the validity of the Goldthorpe class schema’, *European Sociological Review*, 8(3), pp. 211–232. Available at: https://doi.org/10.1093/oxfordjournals.esr.a036638.

Evans, K. (2007) ‘Concepts of bounded agency in education, work, and the personal lives of young adults’, *International Journal of Psychology*, 42(2), pp. 85–93. Available at: https://doi.org/10.1080/00207590600991237.

Feinstein, L., Duckworth, K. and Sabates, R. (2004) *A model of the inter-generational transmission of educational success*. London: Centre for Research on the Wider Benefits of Learning, Institute of Education.

Field, F. (2010) *The Foundation Years: preventing poor children becoming poor adults, The report of the Independent Review on Poverty and Life Chances*. The report of the Independent Review on Poverty and Life Chances.

Franklin, M.N. and Page, E.C. (1984) ‘A Critique of the Consumption Cleavage Approach in British Voting Studies’, *Political Studies*, 32(4), pp. 521–536. Available at: https://doi.org/10.1111/j.1467-9248.1984.tb01543.x.

Furlong, A. (2010) ‘Transitions from education to work: new perspectives from Europe and beyond’, *British Journal of Sociology of Education*, 31(4), pp. 515–518. Available at: https://doi.org/10.1080/01425692.2010.484926.

Furlong, A. *et al.* (2017) *Young people in the labour market: Past, present, future.* Routledge.

Furlong, A. and Cartmel, F. (1997) ‘Risk and uncertainty in the youth transition’, *YOUNG*, 5(1), pp. 3–20. Available at: https://doi.org/10.1177/110330889700500102.

Furlong, A. and Cartmel, F. (2006) *Young people and Social Change*. McGraw-Hill Education.

Furlong, A. and Cartmel, F. (2007) *Young people and social change: New perspectives.* Open University Press.

Galindo-Rueda, F. (2003) ‘Employer Learning and Schooling-Related Statistical Discrimination in Britain’, *SSRN Electronic Journal* [Preprint]. Available at: https://doi.org/10.2139/ssrn.412483.

Gayle, V. (1998) ‘“Structural And Cultural Approaches To Youth: Structuration theory and bridging the gap”, Youth and Policy, 61, 59‐72.’

Gayle, V., Lambert, P. and Murray, S. (2009) ‘School-to-Work in the 1990s: Modelling Transitions with Large-Scale Datasets’, in R. Brooks (ed.) *Transitions from Education to Work*. London: Palgrave Macmillan UK, pp. 17–41. Available at: https://doi.org/10.1057/9780230235403\_2.

Gayle, V. and Lambert, P.S. (2009) ‘Logistic Regression Models in Sociological Research’, *DAMES Node, Technical Paper* [Preprint].

Giddens, A. (1979) *Central problems in social theory: Action, structure, and contradiction in social analysis*. Uni of California Press.

Giddens, A. (1989) *Constitution of Society: Outline of the Theory of Structuration*. Polity Press.

Giddens, A. *et al.* (1991) *Introduction to Sociology*. New York: Norton.

Goldthorpe, J.H. (1980) *Social Mobility and Class Structure in Modern Britain*. Clarendon.

Goldthorpe, J.H. (1998) ‘Rational Action Theory for Sociology’, *The British Journal of Sociology*, 49(2), p. 167. Available at: https://doi.org/10.2307/591308.

Goldthorpe, J.H. and Hope, K. (1974) ‘The social grading of occupations: A new approach and scale.’

Goldthorpe, J.H. and Marshall, G. (1992) ‘The promising future of class analysis: a response to recent critiques’.

Goodwin, J. *et al.* (2020) ‘Returning to YTS: the long-term impact of youth training scheme participation’, *Journal of Youth Studies*, 23(1), pp. 28–43. Available at: https://doi.org/10.1080/13676261.2019.1710484.

Goodwin, J. and O’Connor, H. (2005) ‘Exploring Complex Transitions: Looking Back at the “Golden Age” of From School to Work’, *Sociology*, 39(2), pp. 201–220. Available at: https://doi.org/10.1177/0038038505050535.

Gow, D. (1987) *26 Jan 1987, page 3 - The Guardian at The Guardian*, *Newspapers.com*. Available at: https://theguardian.newspapers.com/image/259908894/ (Accessed: 5 July 2024).

Gregg, P. (2001) ‘The Impact of Youth Unemployment on Adult Unemployment in the NCDS’, *The Economic Journal*, 111(475), pp. F626–F653. Available at: https://doi.org/10.1111/1468-0297.00666.

Gregg, P. (2012) ‘Occupational Coding for the National Child Development Study (1969, 1991-2008) and the 1970 British Cohort Study (1980, 2000-2008).’, *CLS Cohort Studies* [Preprint]. Available at: https://doi.org/10.5255/UKDA-SN-7023-1.

Gregg, P. and Tominey, E. (2005) ‘The wage scar from male youth unemployment’, *Labour Economics*, 12(4), pp. 487–509. Available at: https://doi.org/10.1016/j.labeco.2005.05.004.

Grusky, D. (1994) ‘The Contours of Social Stratification’, in *Social Stratification in Sociological persepective*. Westvoew Press.

Guinea-Martin, D. and Elliott, J. (2008) ‘Economic position and occupational segregation in the 1990s: A comparison of the ONS Longitudinal Study and the 1958 National Child Development Study’, *CLS Cohort Studies* [Preprint].

Hamnett, C., McDowell, L. and Sarre, P. (1989) *Restructuring Britain: The changing social structure*. SAGE.

Hamnett, C. and Mullings, B. (1992) ‘A New Consumption Cleavage? The Case of Residential Care for the Elderly’, *Environment and Planning A: Economy and Space*, 24(6), pp. 807–820. Available at: https://doi.org/10.1068/a240807.

Hancock, M. and Peters, A. (2021) ‘1970 British Cohort Study, Activity Histories (1986 - 2016)’, *UCL Centre for Longitudinal Studies* [Preprint].

Hardt, J. *et al.* (2013) ‘Multiple Imputation of Missing Data: A Simulation Study on a Binary Response’, *Open Journal of Statistics*, 03(05), pp. 370–378. Available at: https://doi.org/10.4236/ojs.2013.35043.

Hawkes, D. and Plewis, I. (2006) ‘Modelling non-response in the National Child Development Study’, *Journal of the Royal Statistical Society*, 169(3), pp. 479–491. Available at: https://doi.org/10.1111/j.1467-985X.2006.00401.x.

Healy, K. (1998) ‘Conceptualising Constraint: Mouzelis, Archer and the Concept of Social Structure’, 32(3).

Hitlin, S. and Elder, G.H. (2007) ‘Time, Self, and the Curiously Abstract Concept of Agency\*’, *Sociological Theory*, 25(2), pp. 170–191. Available at: https://doi.org/10.1111/j.1467-9558.2007.00303.x.

Hitlin, S. and Johnson, M.K. (2015) ‘Reconceptualizing Agency within the Life Course: The Power of Looking Ahead.’, *American Journal of Sociology*, 120(5), pp. 1429–1472. Available at: https://doi.org/10.1086/681216.

Holm, A. and Jæger, M.M. (2011) ‘Dealing with selection bias in educational transition models: The bivariate probit selection model’, *Research in Social Stratification and Mobility*, 29(3), pp. 311–322. Available at: https://doi.org/10.1016/j.rssm.2011.02.002.

HomeOwners Alliance (2012) ‘The death of a dream: the crisis in homeownership in the UK’. HomeOwners Alliance Report.

*Housing Act 1980* (1980). Available at: https://www.legislation.gov.uk/ukpga/1980/51 (Accessed: 23 May 2024).

Howieson, C. and Iannelli, C. (2008) ‘The effects of low attainment on young people’s outcomes at age 22-23 in Scotland’, *British Educational Research Journal*, 34(2), pp. 269–290. Available at: https://doi.org/10.1080/01411920701532137.

Hu, B., Shao, J. and Palta, M. (2006) ‘PSEUDO-R2 IN LOGISTIC REGRESSION MODEL’, *Statistica Sinica*, 16, pp. 847–860.

Hutchison, D., Prosser, H. and Wedge, P. (1979) ‘The Prediction of Educational Failure’, *Educational Studies*, 5(1), pp. 73–82. Available at: https://doi.org/10.1080/0305569790050109.

Hyuk Lee, J. and Huber Jr., J.C. (2021) ‘Evaluation of Multiple Imputation with Large Proportions of Missing Data: How Much Is Too Much?’, *Iranian Journal of Public Health* [Preprint]. Available at: https://doi.org/10.18502/ijph.v50i7.6626.

Iannelli, C. and Smyth, E. (2017) ‘Curriculum choices and school-to-work transitions among upper-secondary school leavers in Scotland and Ireland’, *Journal of Education and Work*, 30(7), pp. 731–740. Available at: https://doi.org/10.1080/13639080.2017.1383093.

‘Industrial Training: Government Proposals’ (1962) *Government White Paper* [Preprint].

ISER (2024) *Changes in the final version*. Available at: https://www.iser.essex.ac.uk/archives/nssec/changes-in-the-final-version (Accessed: 8 January 2024).

Johnson, D.R. and Young, R. (2011) ‘Toward Best Practices in Analyzing Datasets with Missing Data: Comparisons and Recommendations’, *Journal of Marriage and Family*, 73(5), pp. 926–945. Available at: https://doi.org/10.1111/j.1741-3737.2011.00861.x.

Jones, G. (1986) ‘Youth in the social structure: transitions to adulthood and their stratification by class and gender’, *PhD thesis* [Preprint].

Jones, K. (2016) *Education in Britain: 1944 to the present.* John Wiley & Sons.

Jones, M.P. (1996) ‘Indicator and Stratification Methods for Missing Explanatory Variables in Multiple Linear Regression’, *Journal of the American Statistical Association*, 91(433).

Karlson, K.B. and Holm, A. (2011) ‘Decomposing primary and secondary effects: A new decomposition method’, *Research in Social Stratification and Mobility*, 29(2), pp. 221–237. Available at: https://doi.org/10.1016/j.rssm.2010.12.005.

Karlson, K.B., Holm, A. and Breen, R. (2012) ‘Comparing Regression Coefficients Between Same-sample Nested Models Using Logit and Probit: A New Method’, *Sociological Methodology*, 42(1), pp. 286–313. Available at: https://doi.org/10.1177/0081175012444861.

Kogan, M. (2006) ‘Anthony Crosland: intellectual and politician’, *Oxford Review of Education*, 32(1), pp. 71–86. Available at: https://doi.org/10.1080/03054980500496452.

Kohler, U., Karlson, K.B. and Holm, A. (2011) ‘Comparing Coefficients of Nested Nonlinear Probability Models’, *The Stata Journal: Promoting communications on statistics and Stata*, 11(3), pp. 420–438. Available at: https://doi.org/10.1177/1536867X1101100306.

Lambert, P. and Barnett, C. (2021) ‘Optimising the use of measures of social stratification in research with intersectional and longitudinal analytical priorities’, in Nico, M. and Pollock, G., *The Routledge Handbook of Contemporary Inequalities and the Life Course*. 1st edn. London: Routledge, pp. 188–198. Available at: https://doi.org/10.4324/9780429470059-18.

Lambert, P.S. *et al.* (2008) ‘The importance of specificity in occupation‐based social classifications’, *International Journal of Sociology and Social Policy*. Edited by R.M. Blackburn, 28(5/6), pp. 179–192. Available at: https://doi.org/10.1108/01443330810881231.

Lasko, R. (1978) ‘The Work Experience Programme.’, *Department of Employment Gazette*, 86(3), pp. 294–297.

Lekfuangfu, W.N. and Lordan, G. (2022) ‘Documenting occupational sorting by gender in the UK across three cohorts: does a grand convergence rely on societal movements?’, *Empirical Economics* [Preprint]. Available at: https://doi.org/10.1007/s00181-022-02314-5.

Leuze, K. (2010) *Smooth Path or Long and Winding Road? How Institutions Shape the Transition from Higher Education to Work*. Budrich UniPress. Available at: https://doi.org/10.3224/94075542.

Lindley, R.M. (1996) ‘The school-to-work transition in the United Kingdom’, *International Labour Review*, p. 23.

Little, R.J., Carpenter, J.R. and Lee, K.J. (2022) ‘A Comparison of Three Popular Methods for Handling Missing Data: Complete-Case Analysis, Inverse Probability Weighting, and Multiple Imputation’, *Sociological Methods & Research*, p. 004912412211138. Available at: https://doi.org/10.1177/00491241221113873.

Little, R.J. and Rubin, D.B. (2019) *Statistical analysis with missing data*. John Wiley & Sons.

Lynch, J. and Von Hippel, P.T. (2013) ‘Efficiency Gains from Using Auxiliary Variables in Imputation’, *Cornell University Library* [Preprint].

Lynn, P. and Kaminska, O. (2010) ‘Weighting Strategy for Understanding Society’, *Institute for Social and Economic Research* [Preprint].

Lyotard, J.F. (1984) *The postmodern condition: A report on knowledge*. U of Minnesota Press.

Machin, S. and Vignoles, A. (2005) ‘Educational inequality: the widening socio-economic gap’, *Fiscal Studies*, 25(2), pp. 107–128. Available at: https://doi.org/10.1111/j.1475-5890.2004.tb00099.x.

Mackinnon, N. (2001) ‘Labour Market Trends July 2001’, *Labour Market Trends* [Preprint].

Maclure, S. (1978) *Education and Youth Employment in Great Britain*. ERIC.

Madley-Dowd, P. *et al.* (2019) ‘The proportion of missing data should not be used to guide decisions on multiple imputation’, *Journal of Clinical Epidemiology*, 110, pp. 63–73. Available at: https://doi.org/10.1016/j.jclinepi.2019.02.016.

Makeham, P. (1980) ‘Youth Unemployment, Research Paper’, *Department of Employment*, 10.

Makepeace, G., Dolton, P. and Joshi, H. (2004) ‘Gender earnings differentials across individuals over time in British cohort studies’, *International Journal of Manpower*, 25(3/4), pp. 251–263. Available at: https://doi.org/10.1108/01437720410541380.

Marsden, P.V. and Wright, J.D. (eds) (2010) *Handbook of survey research*. Second edition. Bingley: Emerald Group Publ.

Martin, P., Schoon, I. and Ross, A. (2008) ‘Beyond Transitions: Applying Optimal Matching Analysis to Life Course Research’, *International Journal of Social Research Methodology*, 11(3), pp. 179–199. Available at: https://doi.org/10.1080/13645570701622025.

Mayer, K.U. (2004) ‘Whose Lives? How History, Societies, and Institutions Define and Shape Life Courses’, *Research in Human Development*, 1(3), pp. 161–187. Available at: https://doi.org/10.1207/s15427617rhd0103\_3.

Mayer, K.U. (2009) ‘New Directions in Life Course Research’, *Annual Review of Sociology*, 35(1), pp. 413–433. Available at: https://doi.org/10.1146/annurev.soc.34.040507.134619.

Mayer, K.U. and Schoepflin, U. (2022) ‘The State and the Life Course’, p. 24.

McFadden, D. (1972) ‘Conditional logit analysis of qualitative choice behavior.’, in *Frontiers in Econometrics*. Academic Press.

Micklewright, J. (1989) ‘Choice at Sixteen’, *Economica*, 56(221), pp. 25–39. Available at: https://doi.org/10.2307/2554492.

Mostafa, T. and Wiggins, R.D. (2014) ‘Handling attrition and non-response in the 1970 British Cohort Study’, *CLS Working Paper* [Preprint].

Mouzelis, N. (1997) ‘Social and System Integration: Lockwood, Habermas, Giddens’, *Sociology*, 31(1).

Murray, C. (2012) ‘Flagship work programme a “miserable failure”’, *Reuters*, 28 November. Available at: https://www.reuters.com/article/economy/flagship-work-programme-a-miserable-failure-idUSLNE8AQ00N/ (Accessed: 6 July 2024).

Murray, S. and Gayle, V. (2012) ‘Youth Transitions’.

Murray, S.J. (2011) ‘Growing up in the 1990s: Tracks and trajectories of the “Rising 16’s”: A longitudinal analysis using the British Household Panel Survey.’, p. 354.

Nagelkerke, N.J. (1991) ‘A note on a general definition of the coefficient of determination’, *Biometrika*, 78(3), pp. 691–692.

National Children’s Bureau (1981) ‘ncds4\_1981\_part\_1\_data\_dictionary\_questionnaires\_showcards’, *National Children’s Bureau* [Preprint].

Neath, A.A. and Cavanaugh, J.E. (2012) ‘The Bayesian information criterion: background, derivation, and applications’, *WIREs Computational Statistics*, 4(2), pp. 199–203. Available at: https://doi.org/10.1002/wics.199.

*Newsom Report* (1963). Available at: http://www.educationengland.org.uk/documents/newsom/newsom1963.html (Accessed: 10 November 2022).

Neyt, B. *et al.* (2018) ‘Does Student Work Really Affect Educational Outcomes? A Review of the Literature’, *Journal of Economic Surveys* [Preprint].

Norton, E.C. and Dowd, B.E. (2018) ‘Log Odds and the Interpretation of Logit Models’, *Health Services Research*, 53(2), pp. 859–878. Available at: https://doi.org/10.1111/1475-6773.12712.

Office for National Statistics (2014) ‘Young people in the labour market’, *ONS* [Preprint].

Olle, H. (2022) ‘The New Deal for Young People (NDYP)’, *Edge Foundation* [Preprint].

ONS (2023) *Unemployment rate*. Available at: https://www.ons.gov.uk/employmentandlabourmarket/peoplenotinwork/unemployment/timeseries/mgsx/lms (Accessed: 1 May 2023).

Orchard, D.W.J. (1970) ‘DEP direct training services’, *Employment and Productivity Gazette*, October, pp. 856–61.

Pakulski, J. and Waters, M. (1996) *The death of class*. SAGE.

Parker, H.M.D. (1957) *Manpower: a study of war-time policy and administration*. HM Stationary Office.

Parsons, S., Green, F. and Wiggins, D. (2016) ‘Higher Education and Occupational Returns: do returns vary according to students’ social origins?’, *Centre for Longitudinal Studies* [Preprint].

Payne, J. (1987) ‘Unemployment, Apprenticeships and Training: does it pay to stay on at school?’, *British Journal of Sociology of Education*, 8(4), pp. 425–445. Available at: https://doi.org/10.1080/0142569870080405.

Payne, J. (1995) ‘Routes beyond compulsory schooling’, *Policy Studies Institute*, p. 98.

Pearson qualifications (2023a) *About CSEs*. Available at: https://qualifications.pearson.com/en/support/support-topics/understanding-our-qualifications/our-qualifications-explained/about-cses.html (Accessed: 9 May 2023).

Pearson qualifications (2023b) *About O levels*. Available at: https://qualifications.pearson.com/en/support/support-topics/understanding-our-qualifications/our-qualifications-explained/about-o-levels.html (Accessed: 8 May 2023).

Perry, P.J.C. (1976) *The evolution of British manpower policy: from the Statute of Artificers 1563 to the Industrial Training Act 1964*. British Association for Commercial and industrial Education.

Power, C. and Elliott, J. (2006) ‘Cohort profile: 1958 British birth cohort (National Child Development Study)’, *International Journal of Epidemiology*, 35(1), pp. 34–41. Available at: https://doi.org/10.1093/ije/dyi183.

Prandy, K. (1990) ‘The Revised Cambridge Scale of Occupations’, *Sociology*, 24(4), pp. 629–655.

Prandy, K. (1999) ‘Class, stratification and inequalities in health: a comparison of the Registrar‐General’s Social Classes and the Cambridge Scale’, *Sociology of Health & Illness*, 21(4), pp. 466–484. Available at: https://doi.org/10.1111/1467-9566.00167.

Prandy, K. and Lambert, P. (2003) ‘Marriage, Social Distance and the Social Space:: An Alternative Derivation and Validation of the Cambridge Scale’, *Sociology*, 37(3), pp. 397–411. Available at: https://doi.org/10.1177/00380385030373001.

Profillidis, V.A. and Botzoris, G.N. (2019) ‘Trend Projection and Time Series Methods’, in *Modeling of Transport Demand*. Elsevier, pp. 225–270. Available at: https://doi.org/10.1016/B978-0-12-811513-8.00006-6.

Raffe, D. (1984) ‘The Transition from School to Work and the Recession: evidence from the Scottish School Leavers Surveys, 1977‐1983’, *British Journal of Sociology of Education*, 5(3), pp. 247–265. Available at: https://doi.org/10.1080/0142569840050303.

*Robbins Report* (1963). Available at: http://www.educationengland.org.uk/documents/robbins/robbins1963.html (Accessed: 28 November 2022).

Roberts, K. (2003) ‘Change and Continuity in Youth Transitions in Eastern Europe: Lessons for Western Sociology’, *The Sociological Review*, 51(4), pp. 484–505. Available at: https://doi.org/10.1111/j.1467-954X.2003.00432.x.

Roberts, K., Dench, S. and Richardson, D. (1987) ‘The changing structure of youth labour markets’, *Department of Employment*, 59.

Roberts, K.S., Clark, C. and Wallace, C. (1994) ‘Flexibility and Individualisaton: A Comparison of Transitions into Employment in England and Germany’, *Sociology*, 28(1). Available at: https://doi.org/10.1177/0038038594028001004.

Rose, D. and Pevalin, D.J. (2001) ‘The National Statistics Socio-economic Classification: Unifying Official and Sociological Approaches to the Conceptualisation and Measurement of Social Class’, *ISER Working Papers* [Preprint].

Rose, D. and Pevalin, D.J. (2002) ‘The National Statistics Socio-economic Classification: Unifying Official and Sociological Approaches to the Conceptualisation and Measurement of Social Class’, *Sociétés contemporaines* [Preprint].

Rose, P. and Pevalin, D. (2010) *Standard occupational classification 2010*. Basingstoke, Hampshire: Palgrave Macmillan.

Ryan, P. and Unwin, L. (2001) ‘Apprenticeship in the British “Training Market”’, *National Institute Economic Review*, 178, pp. 99–114. Available at: https://doi.org/10.1177/002795010117800114.

Saunders, P. (2003) *Social Theory and the Urban Question*. Routledge.

Saunders, P. (2021) *A Nation of Home Owners*. Routledge.

Savage, L. (2011) ‘Snakes and Ladders: who climbs the rungs of the earnings ladder’, *Resolution Foundation* [Preprint].

Savage, M. and Egerton, M. (1997) ‘Social Mobility, Individual Ability and the Inheritance of Class Inequality’, *Sociology*, 31(4), pp. 645–672. Available at: https://doi.org/10.1177/0038038597031004002.

Schmitt, C. (2021) ‘The impact of economic uncertainty, precarious employment, and risk attitudes on the transition to parenthood’, *Advances in Life Course Research*, 47, p. 100402. Available at: https://doi.org/10.1016/j.alcr.2021.100402.

Schoon, I. *et al.* (2001) ‘Transitions from school to work in a changing social context’, *YOUNG*, 9(1), pp. 4–22. Available at: https://doi.org/10.1177/110330880100900102.

Schoon, I. (2007) ‘Adaptations to changing times: Agency in context’, *International Journal of Psychology*, 42(2), pp. 94–101. Available at: https://doi.org/10.1080/00207590600991252.

Schoon, I. (2010) ‘Becoming Adult: The Persisting Importance of Class and Gender’, in Scott, J., Crompton, R., and Lyonette, C., *Gender Inequalities in the 21st Century*. Edward Elgar Publishing, p. 13500. Available at: https://doi.org/10.4337/9781849805568.00008.

Schoon, I. (2012) ‘Planning for the Future in Times of Social Change’, *Child Development Perspectives*, p. n/a-n/a. Available at: https://doi.org/10.1111/cdep.12003.

Schoon, I. (2020) ‘Navigating an Uncertain Labor Market in the UK: The Role of Structure and Agency in the Transition from School to Work’, *The ANNALS of the American Academy of Political and Social Science*, 688(1), pp. 77–92. Available at: https://doi.org/10.1177/0002716220905569.

Schoon, I. (2022) ‘Planning for the Future: Changing Education Expectations in Three British Cohorts’, p. 22.

Schoon, I., Martin, P. and Ross, A. (2007) ‘Career transitions in times of social change. His and her story’, *Journal of Vocational Behavior*, 70(1), pp. 78–96. Available at: https://doi.org/10.1016/j.jvb.2006.04.009.

Schoon, I., Ross, A. and Martin, P. (2009) ‘Sequences, patterns, and variations in the assumption of work and family-related roles: evidence from two British birth cohorts’, in *Transitions from school to work: Globalization, individualization, and patterns of diversity*. Cambridge University Press.

Scott, J. and Freese, J. (2001) ‘FITSTAT: Stata module to compute fit statistics for single equation regression models’, *Statistical Software Components S407201* [Preprint].

Seaman, S.R. *et al.* (2012) ‘Combining Multiple Imputation and Inverse‐Probability Weighting’, *Biometrics*, 68(1), pp. 129–137. Available at: https://doi.org/10.1111/j.1541-0420.2011.01666.x.

Seaman, S.R. and White, I.R. (2013) ‘Review of inverse probability weighting for dealing with missing data’, *Statistical Methods in Medical Research*, 22(3), pp. 278–295. Available at: https://doi.org/10.1177/0962280210395740.

Shanahan, M.J. (2000) ‘Pathways to Adulthood in Changing Societies: Variability and Mechanisms in Life Course Perspective’, *Review of Sociology*, 26(1), pp. 667–692. Available at: https://doi.org/10.1146/annurev.soc.26.1.667.

Shepherd, P. (1995) ‘The National Child Development Study (NCDS)’.

Sianesi, B., Dearden, L. and Blundell, R. (2003) *Evaluating the impact of education on earnings in the UK: Models, methods and results from the NCDS*. Working Paper Series. IFS. Available at: https://doi.org/10.1920/wp.ifs.2003.0320.

Silverwood, R. *et al.* (2021) ‘Handling missing data in the National Child Development Study: User guide (Version 2).’

Smith, I. (1997) ‘Explaining the Growth of Divorce in Great Britain’, *Scottish Journal of Political Economy*, 44(5), pp. 519–543. Available at: https://doi.org/10.1111/1467-9485.00073.

Smith, T.J. and McKenna, C.M. (2013) ‘A Comparison of Logistic Regression Pseudo R2 Indices’, 39.

Smithers, A. (1993) ‘All Our Futures: Britain’s Education Revolution’, *A Dispatches Report on Education.* [Preprint].

*SOC 2000 - Office for National Statistics* (2000). Available at: https://www.ons.gov.uk/methodology/classificationsandstandards/standardoccupationalclassificationsoc/socarchive (Accessed: 8 January 2024).

Spence, A. (2011) ‘Labour market’, *Social Trends*, 41, pp. 212–236.

*Spens Report* (1938). Available at: http://www.educationengland.org.uk/documents/spens/spens1938.html (Accessed: 10 November 2022).

Steiner, R., Hirschi, A. and Akkermans, J. (2021) ‘Many Roads Lead to Rome: Researching Antecedents and Outcomes of Contemporary School-To-Work Transitions’, *Journal of Career Development*, p. 089484532110635. Available at: https://doi.org/10.1177/08948453211063580.

Stevenson (1913) ‘Annual report for the 1911 of the registrar-general’, 182(4708), pp. 1491–1492. Available at: https://doi.org/10.1016/S0140-6736(01)78008-7.

Stevenson, T.H.C. (1928) ‘The Vital Statistics of Wealth and Poverty’, *Journal of the Royal Statistical Society*, 91(2), p. 207. Available at: https://doi.org/10.2307/2341530.

Stopforth, S. (2020) ‘Parental Socio-Economic Background and Children’s School-Level GCSE Attainment’, *PhD thesis*, p. 306.

Szreter, S.R.S. (1984) ‘The Genesis of the Registrar-General’s Social Classification of Occupations’, *The British Journal of Sociology*, 35(4), p. 522. Available at: https://doi.org/10.2307/590433.

Taylor, M.F.E. *et al.* (2018) ‘British Household Panel Survey User Manual Volume A: Introduction, Technical Report and Appendices’. [object Object]. Available at: https://doi.org/10.5255/UKDA-SN-5151-2.

Tjur, T. (2009) ‘Coefficients of Determination in Logistic Regression Models—A New Proposal: The Coefficient of Discrimination’, *The American Statistician*, 63(4), pp. 366–372. Available at: https://doi.org/10.1198/tast.2009.08210.

Trades Union Congress (2013) ‘The UK’s low pay recovery’, *London: TUC* [Preprint].

Treiman, D.J. (1977) ‘Developing the Scale’, in *Occupational Prestige in Comparative Perspective*. Elsevier, pp. 159–189. Available at: https://doi.org/10.1016/B978-0-12-698750-8.50016-4.

Uberti, L.J. (2022) ‘Interpreting logit models’, *The Stata Journal: Promoting communications on statistics and Stata*, 22(1), pp. 60–76. Available at: https://doi.org/10.1177/1536867X221083855.

University College London (2023) ‘1970 British Cohort Study: Activity Histories, 1986-2016’. UCL Institute of Education, Centre for Longitudinal Studies. Available at: https://doi.org/10.5255/UKDA-SN-6943-4.

University of London (2023a) ‘1970 British Cohort Study: Age 29, Sweep 6, 1999-2000’,. Institute of Education, Centre for Longitudinal Studies. Available at: https://doi.org/10.5255/UKDA-SN-5558-3.

University of London (2023b) ‘National Child Development Study: Age 23, Sweep 4, 1981, and Public Examination Results, 1978’. Institute of Education, Centre for Longitudinal Studies. Available at: https://doi.org/10.5255/UKDA-SN-5566-1.

University of London (2023c) ‘National Child Development Study: Childhood Data from Birth to Age 16, Sweeps 0-3, 1958-1974’. Institute of Education, Centre for Longitudinal Studies. Available at: https://doi.org/10.5255/UKDA-SN-5565-2.

University of London, Butler, N. and Bynner, J. (2023) ‘1970 British Cohort Study: Age 10, Sweep 3, 1980’. Institute of Education, Centre for Longitudinal Studies. Available at: https://doi.org/10.5255/UKDA-SN-3723-8.

University of London and Bynner, J. (2024) ‘1970 British Cohort Study: Age 26, Sweep 5, 1996.’ Institute of Education, Centre for Longitudinal Studies. Available at: https://doi.org/10.5255/UKDA-SN-3833-3.

Vickerstaff, S.A. (2003) ‘Apprenticeship in the `Golden Age’: Were Youth Transitions Really Smooth and Unproblematic Back Then?’, *Work, Employment and Society*, 17(2), pp. 269–287. Available at: https://doi.org/10.1177/0950017003017002003.

Von Hippel, P.T. (2009) ‘How to Impute Interactions, Squares, and Other Transformed Variables’, *Sociological Methodology*, 39(1), pp. 265–291. Available at: https://doi.org/10.1111/j.1467-9531.2009.01215.x.

Wallace, C. and Cross, M. (1990) *Youth in Transition: the sociology of youth and youth policy*. Psychology Press.

Westoff, C.F. and Ryder, N. (2015) *The contraceptive revolution.* Princeton University Press.

White, I.R., Royston, P. and Wood, A.M. (2011) ‘Multiple imputation using chained equations: Issues and guidance for practice’, *Statistics in Medicine*, 30(4), pp. 377–399. Available at: https://doi.org/10.1002/sim.4067.

Williams, M. (2017) ‘An old model of social class? Job characteristics and the NS-SEC schema’, *Work, Employment and Society*, 31(1), pp. 153–165. Available at: https://doi.org/10.1177/0950017016653087.

Wulff, J.N. (2015) ‘Interpreting Results From the Multinomial Logit Model: Demonstrated by Foreign Market Entry’, *Organizational Research Methods*, 18(2), pp. 300–325. Available at: https://doi.org/10.1177/1094428114560024.

Young, R. and Johnson, D.R. (2011) ‘Imputing the Missing Y’s: Implications for Survey Producers and Survey Users’, *Proceedings of the AAPOR Conference Abstracts* [Preprint].

1. Accessed using the UK Data Service, unique identifies: SN5565 and SN5566 [↑](#footnote-ref-1)
2. [↑](#footnote-ref-2)
3. Accessed using the UK Data Service, unique identifies: SN3723, SN3535, SN4751, SN3833, SN5558, SN6943 [↑](#footnote-ref-3)
4. Other includes those respondents that cannot be accurately traced through any of the categories. [↑](#footnote-ref-4)
5. Percentages are based on the participants divided by total cohort. [↑](#footnote-ref-5)
6. The reason sweep 3 has higher participant numbers than sweep 2 etc is due to the way tracking and sampling was handled. Across the BCS, difference organisations took control over this aspect of the survey. [↑](#footnote-ref-6)
7. Age 26 was the first time the cohort member themselves were in complete control of answering the survey itself [↑](#footnote-ref-7)
8. University of Essex, Institute for Social and Economic Research. (2023). *Understanding Society: Waves 1-13, 2009-2022 and Harmonised BHPS: Waves 1-18, 1991-2009: Special Licence Access*. [data collection]. *17th Edition.*UK Data Service. SN: 6931, [DOI: http://doi.org/10.5255/UKDA-SN-6931-16](http://doi.org/10.5255/UKDA-SN-6931-16) [↑](#footnote-ref-8)
9. Defined as non-traditional in this thesis as counter to the ‘traditional’ route to continuing education of going from O’levels straight into sixth-form college to take A’levels to then go to university. Non-traditional in this context means individuals that continued education or schooling in some format but did not go to a sixth form college to take A’levels etc. For this thesis non-traditional and non-academic are synonymous. [↑](#footnote-ref-9)
10. This latter category can be considered an ‘Other’ category. [↑](#footnote-ref-10)
11. ‘’Rule6’’ means N/A. [↑](#footnote-ref-11)
12. Either in Social Housing or privately rented accommodation. [↑](#footnote-ref-12)
13. Appendix 1, table 6.1.5 onwards [↑](#footnote-ref-13)
14. In the case of his 1928 paper Stevenson was primarily focused with assessing the relative strength of class in understanding the phenomena of mortality rates – he found that a class-based approach was a much better approach compared to a study of income or wealth. [↑](#footnote-ref-14)
15. The individual cohorts will themselves be analysed in more granular detail in latter chapters. For now, the focus is on the overall story of youth transitions and the impact, if any, or cohort level change on the story of those transitions. [↑](#footnote-ref-15)
16. For the initial pooled analysis, the SOC 2000 construction of NS-SEC will be used as the primary social stratification variable. In subsequent detailed analysis, other social stratification measures will be assessed in NS-SECs place in the form of a sensitivity analysis. [↑](#footnote-ref-16)
17. Unfortunately, the qv command and subsequent graphing subcommands do not currently work with the multinominal logistic regression models in this chapter – QV estimates are only produced for the first category in the categorical outcome variable, nor does it work with the sub-command ‘’ib().” that is used to identify a specific reference category of a chosen variable such as NS-SEC – this is because the ‘qv’ command predates the implementation of the subcommand ‘’ib().”. The creation of quasi-variance statistics can be completed via a quasi-variance calculator (Firth, 2000). Whilst this does produce the required quasi-variance statistics, there are two notable issues with this direction. The first is that producing quasi-variance statistics outside of Stata breaks the workflow and increases the possibility of manual error. The second is that the given quasi-variance calculator does not provide lower and upper bound 95% CIs for quasi-variance, instead producing a singular quasi-variance statistic. An alternative solution was identified that did not break the workflow and was committed within Stata. The ‘’ib().” The subcommand issue can be overcome by recoding NS-SEC whereby the reference category is first – in this case, recoding NS-SEC 7 as NS-SEC 1 so that Stata is forced to use that category as the reference.

    [↑](#footnote-ref-17)
18. Variable n4118 used [↑](#footnote-ref-18)
19. The variable in question was acatnn236, a categorical variable. [↑](#footnote-ref-19)
20. Burn-in was 20 during imputation. [↑](#footnote-ref-20)
21. Graphs found in appendix. [↑](#footnote-ref-21)
22. There are further combinations of missing observations at variables that are no larger than 3 per cent and are not included in this table. [↑](#footnote-ref-22)